Spring 2013

Real-Time Detection for Scattering Scanning Near-Field Optical Microscopy

Justin Alan Gerber
University of Colorado Boulder

Real-Time Detection for Scattering Scanning Near-Field Optical Microscopy

by

Justin Alan Gerber

A thesis submitted to the

Faculty of the Honors Council of the

University of Colorado in partial fulfillment

of the requirements for the award of

departmental honors in the

Department of Physics

2013

This thesis entitled:
Real-Time Detection for Scattering Scanning Near-Field Optical Microscopy
written by Justin Alan Gerber
has been approved for the Department of Physics

_____
Markus Raschke

_____
Paul, Beale

_____
Gregory, Beylkin

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Gerber, Justin Alan (B.A. Physics)

Real-Time Detection for Scattering Scanning Near-Field Optical Microscopy

Thesis directed by Professor Markus Raschke

Scattering scanning near-field optical microscopy (s-SNOM) is a powerful technique for measuring spectroscopic properties of materials with spatial resolution previously unobtainable due to the diffraction limit. s-SNOM combines scanning probe microscopy (SPM) with spectroscopy to provide sub-diffraction limited spatial resolution information about optical and related properties of matter. Discriminating the weak elastic s-SNOM signal from various undesirable background signals is critical to the success of s-SNOM as a measurement technique. Traditionally this discrimination is achieved through lock-in detection, in which the s-SNOM signal is restricted to, and measured at harmonics of an atomic force microscope tip oscillation frequency Ω. However, this detection technique neglects information at all undetected harmonics. To overcome this loss of information, for my honors thesis, I developed a new s-SNOM detection scheme based upon the real-time acquisition of the s-SNOM signal to ensure no information is lost. With this new detection scheme I have been able to simulate gated detection elastic s-SNOM –a measurement which has not yet been realized experimentally– and determine that it provides no more than background-ridden information readily obtainable with lock-in detection. I have also been able to generate, for the first time, an experimental reconstruction of the distance dependence of the tip-sample interaction for elastic s-SNOM measurements and compare it to theoretical models.

Acknowledgements

I would like to thank Prof. Markus Raschke for providing me with many outstanding opportunities to begin to form my identity as a researcher. These opportunities have included everything from learning the basics of using lab equipment to traveling to Germany for a collaborative project. The skill and experience I've gained under his guidance have prepared me very well for what lies ahead in my career as a researcher.

I would like to particularly thank Sam Berweger for mentoring me in and out of the lab. Following his example has helped me to develop an intuition for experimental physics I never could have learned in a classroom. His guidance on this project has helped me make the transition from fumbling around in the lab figuring out how everything works to pulling all of my experience together to produce results.

It has been a great pleasure working with all of the members of the Raschke group. Their support, feedback, and advice have been critical to success and enjoyment of this project. I am extraordinarily grateful for their willingness to answer my endless slough of questions, scientific and otherwise.

Finally, I would like to thank my friends and family for their endless support and confidence in me. You gave me the confidence in myself I needed to continue to give my all in my studies and my research.

Contents

Chapter

Figures

Figure

Chapter 1

Introduction

1.1 Motivation

Many technologically and scientifically interesting effects arise from material spatial inhomogeneities as well as mesoscopic structures with natural length scales from tens to hundreds of nanometers. Scientists have been motivated to study materials in this regime by such systems as quantum dots [1], nano-transistors [2], single molecule biological samples [3], composite chemical systems [4], topological materials [5] and many others.

To measure matter on these small length scales, it is necessary to have an instrument with spatial resolution finer than the length scale of the systems under consideration. Advanced techniques have been developed to measure the spatial distribution of matter in nanoscopic systems with resolution down to an atomic radius. These techniques include scanning electron microscopy (SEM) [6], scanning tunneling microscopy (STM) [7], atomic force microscopy (AFM) [8], transmission electron microscopy (TEM) [9], and X-ray spectroscopy [10].

These techniques can provide information about the spatial distribution of matter as well as limited information about the electronic properties of the material [11]. It is also possible to study structural, optical, thermal, acoustic and magnetic properties of materials with the techniques mentioned. However, it can be very experimentally difficult and to extract this information using these techniques. On the other hand, researchers have been extraordinarily successful and found relative ease using the interactions between light and matter to study these material properties [12].

How light of a given energy, $\hbar\omega$, interacts with matter depends on fundamental properties of that material such as the electronic density of states [13], physical structure [10], quantum mechanical spin states [14], and many others. Optics –the use of light to study physical systems– is generally much easier to implement experimentally than the aforementioned systems. The ease of optics combined with the vast amounts of information obtainable from it have made optics an indispensable tool for researchers.

Countless techniques utilize light-matter interactions to characterize materials. These techniques include conventional optical microscopy, absorption measurements, fluorescence measurements, and ultra-fast techniques. However, if these techniques are performed –as they conventionally are– using far-field light, such as that produced by lamps and lasers, the best achievable resolution is restricted by a fundamental principle known as the diffraction limit [15]. For optical frequencies this resolution limit is approximately 200 nm.

However, many systems exhibit interesting material properties that vary on length scales smaller than this diffraction limit. These mesoscopic systems are immeasurable with traditional optical spectroscopic techniques. s-SNOM is a relatively new measurement technique which provides sub-diffraction-limited optical spectroscopic imaging. s-SNOM addresses this inability of traditional spectroscopy techniques to provide spectroscopic information on length scales from approximately 1 nm to 100 nm.

s-SNOM is still a relatively new measurement technique. This means it is still in the process of being developed. To this end, various s-SNOM schemes have been developed with their respective advantages and disadvantages in terms of the type of information they provide, the quality of the data (signal-to-noise ratio and contrast), and the ease of measurement.

It is the goal of my honors thesis to develop a new optical detection system for s-SNOM that will provide both more comprehensive optical information than previous implementations as well as a means to investigate the theoretical foundations of s-SNOM.

1.2 Overview

The goal of s-SNOM is to provide spectroscopic information about material surfaces with resolution better than the diffraction limit. In s-SNOM, this is achieved by utilizing an electromagnetic phenomenon

known as the near-field and its interaction with a small probe to perform spectroscopic SPM. SPM refers to a set of techniques whereby a small tip that serves as the probe is placed near to a specific point on a two-dimensional sample, and data is collected from that specific point only. Once this process is repeated at every point on the two-dimensional sample surface, a spatial image of that property can be generated revealing scientifically interesting material information. s-SNOM is a spectroscopic SPM technique. In this section I will give an overview of the techniques and theory around which a s-SNOM system is based.

1.3 Scanning Probe Microscopy

SPM is a technique whereby a small (nm scale) probe –which can be fabricated in a multitude of fashions and from a variety of materials– is brought very close to or into physical contact with the surface of some experimental sample so that it interacts with the sample's local properties. This interaction can be monitored and implemented into a closed-loop feedback system which allows the SPM system to maintain a constant, and small (few nanometers or less) physical separation between the tip and sample. The interaction between the tip and sample can be used to determine the properties of the sample. The tip is raster-scanned over the surface of the sample and the relevant property of the material is recorded at each pixel on the sample and correlated with the tip's position with respect to the sample. With this information, a two-dimensional spatial image of that sample property can be generated. Some properties that can be measured using different SPM techniques are sample topography, conductivity, viscosity, and various electromagnetic parameters [16].

The first scanning probe microscope was the STM [17]. STM utilizes tunneling currents that occur between a conductive tip very close to a conductive sample to moniter sample properties such as conductivity or topography. It was very successful but could only be used with conductive samples and this greatly limited its application.

However, fortuitously, experiments with the STM revealed that the tip felt forces when it was brought very close to the sample. These forces called for the invention of AFM [18]. AFM utilizes these forces felt by the tip to determine the height of the sample. Modern AFM can provide atomic resolution topographic images [8]. However, the resolution of the image is limited by the size of the tip and, in s-SNOM, tips are

typically as large as 20 nm -many times the size of an atom.

For s-SNOM systems, The AFM serves three purposes: holding the tip in the sample interaction region, measuring the height of the sample, and scanning the tip around the sample to generate a two-dimensional map of the topographic and spectroscopic structure of the sample.

1.4 Spectroscopy and the Diffraction Limit

AFM is a valuable technique for measuring sample topography, but provides little information about the sample's underlying optical properties. Spectroscopy encompasses a large class of techniques which determine how matter interacts with light. In spectroscopy, light is shined onto a sample such that it interacts with the sample. Light then either scatters from or passes through the sample and can be analyzed to understand the interaction. This process gives scientists information about the chemical composition of matter [19], the crystal structure of solids [10], ultra-fast quantum transitions [12], semi-conductor properties [20], vibrational characteristics [21] and many other intriguing and technologically interesting characteristics of matter.

Spectroscopy has traditionally been performed by focusing a laser onto a sample with focusing optics and detecting the scattered light again with focusing optics. However, there is a fundamental limit known as the diffraction limit which puts a lower bound on both smallest spatial features which can be resolved by imaging optics as well as the minimum spot size to which a beam of light can be focused.

The first sense of the diffraction limit I have described arises from the fact that when a point source is imaged through a lens with a finite aperture, the light interferes with itself after passing through the aperture and the resultant image is a non-point-like diffraction pattern known as an Airy disk. See Fig. 1.1. The angular separation, θ, between the central maximum and the first minimum of the Airy disk pattern is given by the following expression relating the angular separation to the wavelength of the light, λ, and the diameter of the imaging aperture, D [22].

$$\theta = 1.22\frac{\lambda}{D} \tag{1.1}$$

If two point sources are simultaneously imaged they will each produce an Airy disk. One criteria for

stating that the two points are resolved is if the maximum of one Airy disk falls outside of the minimum of the other Airy disk. This is known as the Rayleigh criterion [23]. Some manipulation of equation 1.1 relates, instead, the spatial separation required to resolve two points, l_{min} to λ and the numerical aperture of the imaging optic, NA.

$$l_{min} = 0.61\frac{\lambda}{\text{NA}} \tag{1.2}$$

The numerical aperture of a lens is defined by $\text{NA} = n\ \sin(\theta)$ where θ is the half angle of the acceptance cone of the imaging optics. Modern optics can achieve numerical apertures up to about 1.4 so it is commonly claimed that the diffraction limit is one half the working wavelength, $\frac{\lambda}{2}$.

The concept of the wavevector is critical to an understanding of spatial resolution in optical (and other) measurement techniques. For a simple plane electromagnetic wave, the wavevector characterizes the direction of propagation as well as the spatial frequency of the wave. It is generally true that a larger wavevector corresponds to larger spatial frequencies and thus higher spatial resolution. Intuitively, this is because the rapidly varying spatial features of an object with large spatial frequencies can used to study objects which vary on larger length scales. In short: the presence of larger wavevectors allows for higher spatial resolution.

The diffraction limit for a lens is arises from the fact that large wavevector components of the fields emitted by the point sources are filtered by free space[1] and the lens itself [24]. Only the low-resolution, small wavevector components are imaged by the lens so the image is smeared.

A similar argument based on the Fourier transformation properties of lenses and, again, the fact that all real optics have finite numerical apertures shows that a beam of light cannot be focused down to a spot smaller than the diffraction limit [24]. This is because the finite aperture of the lens filters out the largest wavevector components of the incident beam so it can only be focused to a finite spot size.

The fact that beam focusing and imaging resolution are both constrained by the diffraction limit imply that any measurement which utilizes far-field light to directly measure spectroscopic properties of a sample can never achieve spatial resolution better than the diffraction limit. This severely impacts scientists ability

[1] More on this in section 1.5

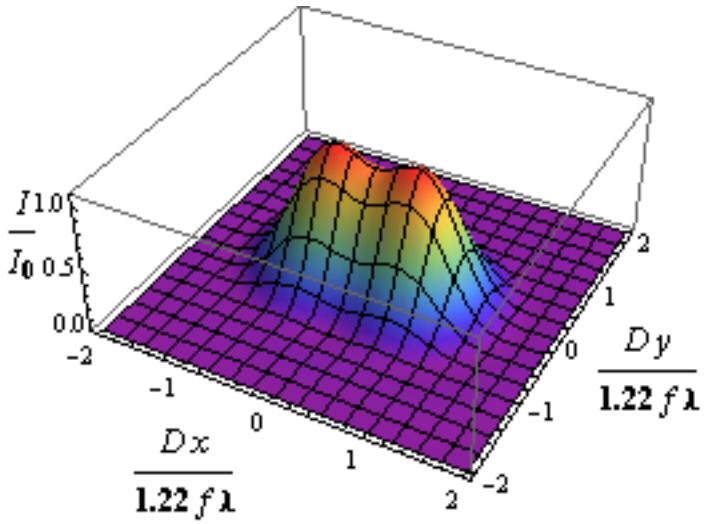


Figure 1.1: Two point sources separated by the Rayleigh criterion distance imaged by a lens. Only a finite range of wavevectors are admitted into the lens so the resultant image cannot have infinite spatial resolution i.e. the points will be smeared into an Airy disks.

to study material properties on interesting nanometer length scales. While it may be possible to use higher-energy light with shorter wavelengths to study smaller samples, this solution is both impractical because of experimental considerations and insufficient due to an energy mismatch between high-energy X-ray photons and vibrational and electronic interactions which have energies in the optical and infrared range. The goal of s-SNOM and other near-field techniques is to measure these and other spectroscopic properties of samples on length scales smaller than the diffraction limit.

1.5 The Near Field

The diffraction limit problem can be cast in the language of wavevectors for solutions to Maxwell's equations. We have the following equations describing the wavevector for familiar propagating plane wave solutions to Maxwell's equations:

$$\begin{aligned} \mathbf{E} &= E_0 e^{i\mathbf{k}\cdot\mathbf{r}} + c.c. \\ \mathbf{k} &= \frac{\omega}{c}\widehat{\mathbf{k}} \\ |\mathbf{k}|^2 &= k_x^2 + k_y^2 + k_z^2 \end{aligned} \tag{1.3}$$

Notice in particular that the magnitude of the wavevector is determined entirely by the frequency of

the light, ω, and the speed of light, c. Also notice that if the components of the wavevector are all real, as in the case of plane waves, then each individual component must be less than the total magnitude of the wavevector. This can be summarized by saying that free space acts as a low-pass momentum-space filter [24]. Since high spatial frequency components of the light are filtered by free space this means it is impossible to use propagating light to image small physical features –a restatement of the diffraction limit.

To circumvent this problem, we must utilize solutions to Maxwell's equations that have spatial frequency components larger than the magnitude of the total wavevector. However, if we have $k_x^2 + k_y^2 > |\mathbf{k}|^2$ then it must also be the case that k_z is imaginary. This introduces a spatially decaying exponential term to the expression for the electric field, $E = E_0 e^{-k_z z}$.

Such decaying solutions to Maxwell's equations arise at the interface between two materials with differing dielectric constants and are known as evanescent waves. Evanescent waves make up, in part, what is known as the near field. The near field is the electromagnetic field that persists in a region within distance $< \lambda$ of a source of light. The presence of high spatial frequency components in the near field and the fact that it only exists very close the material surface imply that the near field can encode high resolution spectroscopic material properties that could be taken advantage of if it were possible to 'probe' the near field.

Chapter 2

Background and Theory

2.1 Near-Field Microscopy

The fundamental idea in near field microscopy is to place a small aperture (smaller than the wavelength of light) very close to the sample surface. This aperture will then interact with the near field of the sample and diffract or scatter the light into the far field where it can be easily detected. It is profitable to think of the tiny aperture as an antenna which converts near-field light in the vicinity of the sample into far-field light which can be detected by traditional means [25].

The idea of using a small aperture to probe the near field and beat the diffraction limit for microscopy dates back to 1928 when Edward Synge corresponded with Albert Einstein on the topic [26]. In 1972, Ash and Nicholls first made a sub-diffraction limited near-field measurement by utilizing an aperture scheme to measure microwave near-field light [27]. Dieter Pohl did the same with optical frequencies in 1984 to obtain spatial resolution of $\frac{\lambda}{20}$ [28]. Pohl made the interesting observation that a doctor's stethoscope of a diameter of about 10 cm, is able to measure sound waves whose wavelengths would be almost 100m –clearly a sub-diffraction limited measurement.

Because of the need for very precise mechanical control of the sample and the probe, near-field optical microscopy techniques were immediately coupled with scanning probe techniques such as AFM and evolved side by side. These techniques came to be known as near-field scanning optical microscopy (NSOM) techniques.

The first NSOM techniques were based on measuring the near field through a tiny aperture and were

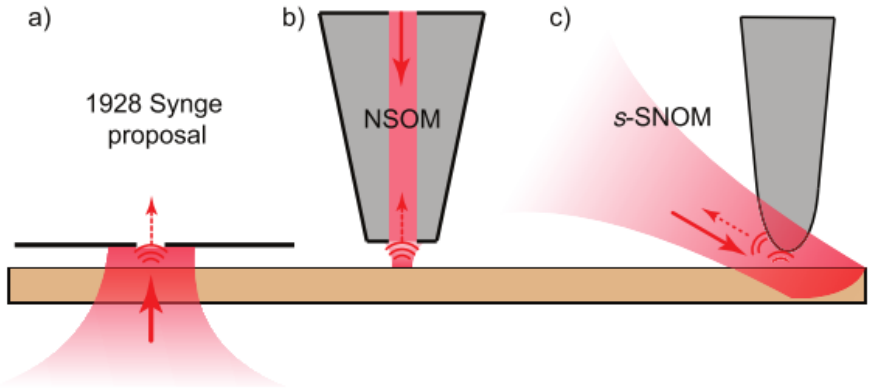


Figure 2.1: a) Synge's original proposal to use a small aperture to collect light emitted from a point with sub-diffraction limited resolution. b) Typical NSOM tip in which light is collected by a waveguide aperture in the center of the tip. c) s-SNOM tip in which the tip scatters near-field light from a localized region into the far field. Figure after Ref. [5].

known as aperture NSOM. In aperture NSOM techniques the probing tip is a small conical structure pointed towards the sample which has an aperture running through its center. The aperture acts as a waveguide into which near field light can be coupled. The probes are commonly constructed by metal coating tiny fiber optics cables. The resolution in aperture NSOM is proportional to the radius of the aperture. However, as the aperture size decreases the amount of light admitted into the aperture decreases as well and thus signal levels drop. This results in a trade-off between signal levels and spatial resolution leading to a practical resolution limit for aperture NSOM of $\frac{\lambda}{10}$ [29].

The first apertureless NSOM systems were demonstrated in the 1990's [30]. Instead of collecting light through an aperture, these NSOM systems utilize light scattered from the tip's apex to convert the near field into a measurable far-field signal. These systems are known as s-SNOM systems. See Fig. 2.1. In s-SNOM, the spatial resolution approximately scales with the tip apex radius. One of the main advantages of s-SNOM over aperture NSOM techniques is that it is possible to fabricate much sharper tips and thus achieve much higher spatial resolution. Furthermore, in certain s-SNOM configurations the tip itself serves to amplify the incident and scattered electric fields leading to higher signal levels [31].

s-SNOM has been utilized to perform various different types of nanometer resolution spectroscopic experiments. It is possible to perform elastic, inelastic, and non-linear s-SNOM measurements. Implemen-

tations include simple elastic scattering experiments to measure dielectric function contrast [32], inelastic tip-enhanced Raman scattering measurements to measure nanometer resolution chemical composition [21], and measurements of spatially resolved second harmonic generation [33] among other applications.

2.2 Modeling the Near-Field Tip-Sample Interaction

The purpose of s-SNOM systems is to use a SPM probe as an antenna to convert the near field in the vicinity of the sample into a detectable optical signal which I will call the scattered near-field signal or just near-field signal. The amplitude and phase of the near field –the quantities of scientific interest– are encoded in the intensity of the scattered near-field signal.[1] To measure spatial contrast in the sample near-field, researchers must have a theoretical model which allows them to infer information about the sample's near field from the far-field scattered near-field signal.

In s-SNOM the scattered near-field intensity depends, to first order, on the dielectric functions of the tip and the sample and the distance between the tip and the sample. Often we are interested in measuring the sample's dielectric function. Because of these parameters, the fundamental prediction of any model for the s-SNOM interaction is the (tip-sample) distance dependence of the near-field signal. I will call this distance dependence a raw approach curve.[2] Examples of approach curves for different theoretical models are shown in Fig. 2.5.

There are various models for the s-SNOM interaction. In this section I will address three such models: The prolate spheroid model, the simple dipole model, and the extended monopole model. As with all theoretical models, these models each have their own limited range of applicability for which they make accurate predictions. One of the experimental techniques I develop in my thesis provides a means to, for the the first time, directly compare these theories to experiment and potentially help to develop new theoretical descriptions. With this in mind, I will now outline the basics of these different theories.

[1] Section 2.5 describes how both amplitude and phase can be encoded and extracted from signal intensities

[2] It is called an approach curve because it is typically measured by slowing approaching the tip towards the sample from a distance. I call it raw to contrast with a demodulated approach curve which I will introduce in section 2.3.

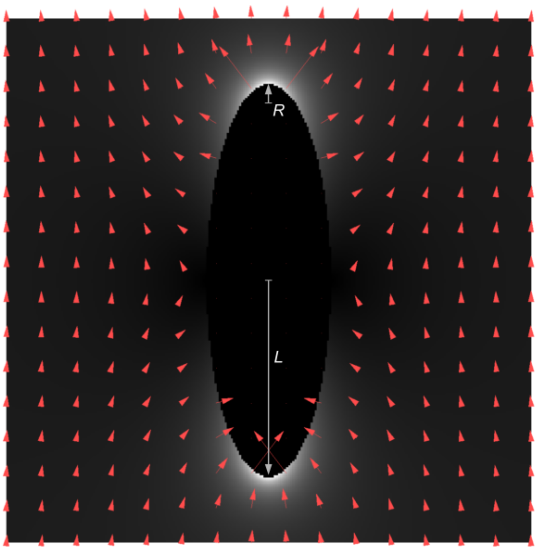


Figure 2.2: Vector plot of the electric field surrounding a conducting spheroid in a constant electric field. Figure after Ref. [34].

2.2.1 Tip Spheroid Model

I briefly mention the tip spheroid model as it is the model to which I will compare the subsequent models. In the tip spheroid model, the tip is modeled as a conducting or dielectric prolate spheroid. Maxwell's equations have been solved in spheroidal geometry so it is not difficult to do electromagnetic calculations in this model. The incident laser field is modeled as a constant, uniform electric field in which the tip spheroid is immersed [35]. In the most basic implementations of this model, the sample is ignored.

The spheroid interacts with this electric field and the result is that the electric field becomes enhanced near the top and bottom of the spheroid. This is because the electric field drives a charge separation in the spheroid and there is a charge buildup at the points of high curvature on the surface which results in a large field enhancement. This effect is known as the lightning rod effect.

The distance dependence of this field enhancement can be expressed as the following [36].

$$E_s(z) = \frac{\frac{2F(L+z)}{z^2+L(2z+R)} + \ln\frac{L-F+z}{L+F+z}}{2\frac{F(L-\epsilon_t R)}{LR(\epsilon_t-1)} - \ln\frac{L-F}{L+F}} E_0 \tag{2.1}$$

F is the distance from the center of the spheroid to a focus. L is the length of the major axis of the spheroid. z is the distance from the apex of the spheroid. ϵ_t is the dielectric constant for the spheroid and

E_0 is the incident electric field. This distance dependence is plotted in Fig. 2.5.

The spheroid models the apex of the tip very well but it does not account for the long range spatial expanse of the tip well. This model is also limited by the fact that an analytical solution when the spheroid is in the presence of a flat dielectric sample is not known. Furthermore, this and all of the following models utilize the quasi-static approximation. This amounts to neglecting the periodic spatial variations in the electric field present in electromagnetic radiation. This approximation renders the theories invalid for distances from the tip larger than the wavelength of the electric field.

2.2.2 Simple Dipole Model

Perhaps the simplest model of the tip-sample interaction is the simple dipole model. In this model, the SPM tip is treated as a dielectric sphere with radius R corresponding to the radius of the tip, and dielectric constant ϵ_{tip} at height d above the sample. The sample is treated as a dielectric plane with dielectric constant ϵ_{surf} [29]. See Fig. 2.3.

A qualitative description of the interaction is as follows. When this system is exposed to an electric field (incident laser), the field excites a dipole in the tip sphere. Because of the dielectric nature of the sample surface, an image dipole is excited in the sample. This coupled dipole system now has an effective polarizability α_{eff}. The oscillating electric field, E_{inc}, interacts with this polarizable system which results in a total electric field with amplitude $E_{tot} \propto \alpha_{eff} E_{inc}$. The coupled dipole system then scatters this electric field and we get $E_{sca} \propto E_{tot}$ scattered into the far field to be detected.

In more detail, it can be seen that α_{eff} depends on the dielectric constants for both the tip and the sample, ϵ_{tip} and ϵ_{surf}, the tip-sample separation, z, and whether the incident electric field is polarized parallel to the tip axis, p-polarized, or perpendicular to the tip axis, s-polarized. This model can be summarized in the following equations [29].

Polarizability of a dielectric sphere:

$$\alpha = 4\pi R^3 \left(\frac{\epsilon_{tip} - 1}{\epsilon_{tip} + 2} \right) \tag{2.2}$$

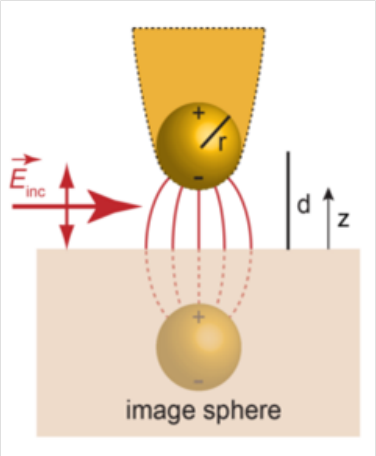


Figure 2.3: In the simple dipole model, the tip is modeled as a polarizable sphere which induces an image sphere in the sample. The interaction of these two dipoles gives rise to an effective polarizability, α, which determines the amplitude of the scattered field. Figure after Ref. [5].

Relative strength of the image dipole:

$$\beta = \frac{\epsilon_{surf} - 1}{\epsilon_{surf} + 1} \tag{2.3}$$

Effective polarizability of sphere coupled to image sphere for p polarized light

$$\alpha_z = \alpha \left(1 - \frac{\alpha\beta}{16\pi(R+z)^3}\right)^{-1} \tag{2.4}$$

Effective polarizability of sphere coupled to image sphere for s polarized light

$$\alpha_x = \alpha \left(1 - \frac{\alpha\beta}{32\pi(R+z)^3}\right)^{-1} \tag{2.5}$$

Notice that the field is enhanced more for p-polarized light than s-polarized light. Because of this, s-SNOM is typically performed with light polarized along the axis of the tip. This makes sense from an antenna standpoint because charges are more free to move in the z direction than the x or y directions. However, there are situations in which s-polarized light is preferable [37].

One of the best features of this model is its simplicity owing to the fact that it is based on simple electricity and magnetism. This makes it flexible and extendable. The dipole model has served researchers well and made successful qualitative predictions about s-SNOM measurements [38].

However, though the dipole model has been successful, it has shortcomings. The biggest shortcoming of the dipole model is that it has very bad quantitative agreement with exact models and experiment. For

example, Fig. 2.5 shows the poor quantitative agreement between the simple dipole and the more exact tip-spheroid model. In the next section I describe a more comprehensive, and also more complicated, model which was later introduced.

2.2.3 Extended Dipole Model

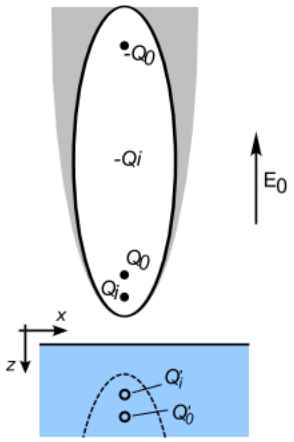


Figure 2.4: In the extended dipole model, the entire tip is now treated as a dipole. The image charges produced by Q_0 interact again with the tip, which is modeled as a spheroid, to produce more charges. The interaction of all of these charges determines an effective polarizability that determines the scattering amplitude. Figure after Ref. [36].

The extended dipole model (sometimes called the monopole model) is similar to the simple dipole model in that it is based on the interaction of an incident electric field interacting with the dielectric functions of the tip and the sample. It differs in how it models the tip and the image charges induced in the sample.

The extended dipole model is motivated by analytical calculations of the electric field surrounding a prolate spheroid (as a model for the tip) illuminated by a constant electric field [34]. The calculations showed that the electric field resembled the field of an extended charge distribution more than the field of a point dipole as in Fig. 2.2 [35].

In the simple dipole model, the effect of the electric field on the tip is to drive a dipole located in center of a sphere inscribed in the tip apex. In the extended dipole model, the electric field, instead, drives a charge separation (similar to a dipole) in which charge Q_0 can be found at the tip apex and charge $-Q_0$ can be found far away from the apex up the tip shaft. Only charges near to the sample surface will participate in the interaction so $-Q_0$ will not play a further role. Q_0 now induces a mirror charge Q_0' in the sample. Now this mirror charge will cause the charges in the tip (modeled as a prolate spheroid) to redistribute. To a very crude approximation, this redistributed charge can be approximated by replacing the spheroid with

another monopole charge Q_i. This induces another mirror charge Q_i' in the sample. Performing calculations on this interaction that capture its recursive nature allow values to be established for Q_0 and Q_i and an effective polarizability to be established.

The resultant expression for the distance dependence of the scattered near-field is as follows [36].

$$\alpha_{eff} = R^2 L \frac{\frac{2L}{R} + \ln\frac{R}{4eL}}{\ln\frac{4L}{e^2}} \left(2 + \frac{\beta(g - \frac{R+z}{L})\ln\frac{4L}{4z+3R}}{\ln\frac{4L}{R} - \beta(g - \frac{3R+4z}{4L})\ln\frac{2L}{2z+R}} \right) \tag{2.6}$$

R is the tip radius, L is the length of the semi-major axis of the spheroid, β is the reflectivity of the sample, g is a parameter related to distribution of the induced electric charges and z is the tip-sample separation.

This theory is advantageous because it is analytical yet able to capture more quantitative features of the interaction than the simple dipole model. However, the expressions themselves are fairly complex and not very intuitively revealing on their own. Furthermore, the extended dipole model introduces more parameters into the model than the simple dipole model so a better fit is expected.

In short, these latter two models I have described here have had success explaining experimental results but, like all theoretical models, are limited. For example, both models assume an infinite dielectric plane for the sample. At a boundary between two dielectrics we expect the models to break down. They also both make major assumptions about the shape of the tip apex which could be very different from reality.

In this thesis one of my main results is a direct measurement of the tip sample interaction as a function of tip-sample separation referred to as a raw optical approach curve. The curve can be directly compared to theoretical predictions to determine the validity of the theory. Hopefully the result described in this thesis will lead to a more comprehensive model of the tip sample interaction.

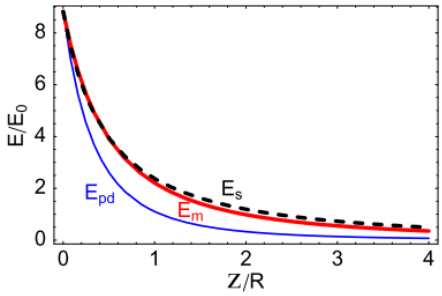


Figure 2.5: Distance dependence of the relative strength of the electric field in the presence of a prolate spheroid dielectric (black). Distance dependence of the scattering from the simple dipole model (blue), and the extended dipole model (red). Figure modified from Ref. [36].

2.3 Lock-in Detection

There are certain SPM implementations which physically oscillate the probe to maintain feedback and a constant tip-sample separation. In tapping mode AFM, for example, the tip is placed, pointing downward, on the end of a cantilever which is driven to oscillate at its resonant frequency. The tip barely makes physical contact with the sample at the bottom of this cantilever oscillation. As seen in the previous section, the scattered near-field signal is strongest when the tip is near the sample and weakest when the tip is far away. Thus, the resultant scattered signal will be an optical signal modulated at frequency Ω. This modulation can be taken advantage of to retrieve additional information about the sample's near field.

The simplest way to think about the signal at the detector, $I(t)$, is as a composition of the time dependence of the tip position, $z(t)$ and distance dependence of the scattered near field signal, $I(z)$. The intensity falling on the detector is $I(z(t))$ where $z(t) = z_0 + A + A\cos(\Omega t)$. A is the tip oscillation amplitude and z_0 is the minimum tip sample separation. $I(z)$ is determined from theoretical models for the near field scattering such as those explained in sections 2.2.2 and 2.2.3.

Since the distance dependence of the near field is a non-linear function of distance we see that $I(z(t))$ has frequency contributions at all integer harmonics, $n\Omega$, of the tip oscillation frequency. This is apparent by looking at the Fourier transform. See Figure 2.6.

Since the signal is restricted to these specific frequencies in Fourier space, it is possible to use a lock-in detector to only detect the signal at a given harmonic. A lock-in detector is an electronic device which takes

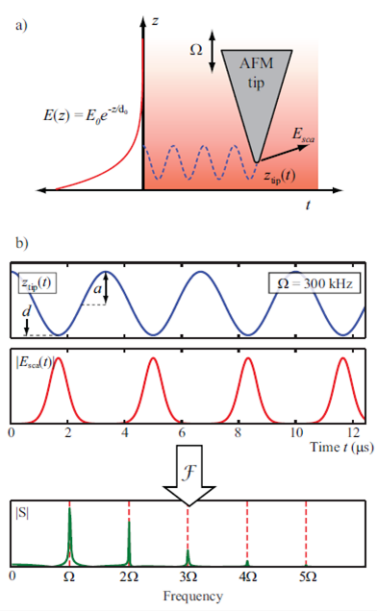


Figure 2.6: a) distance dependence of the near-field scattering, $I(z)$, combined with b) the sinusoidal tip motion, $z(t)$ produces the optical signal $|E_{sca}(t)|$. Taking the Fourier transform of the non-linear signal shows signal peaks at frequencies $n\Omega$. Lock-in detection measures one of these peaks at a time. Figure after Ref. [25].

as an input the electronic signal of interest (optical signal from the MCT in the case of s-SNOM) and a reference signal (tip oscillation signal). The lock-in then the calculates the Fourier component of the signal of interest at the frequency of the reference signal. It is essentially a Fourier filter.

The first advantage to this method is that it eliminates all sources of noise except those present at the lock-in frequency. The next advantage of lock-in detection requires a brief discussion of background signals in s-SNOM. Ideally, in a s-SNOM system the only light exiting the tip-sample interaction region would be near-field light scattered by the tip which contains high spatial resolution information due to the large wavevector components of the near field. However, this is not the case. The incident diffraction limited laser focus is many times larger than the tip apex.[3] Components of this large spot will reflect off of the tip shaft and the sample in unpredictable ways and some of these reflections will be directed towards the detector. This results in a significant undesirable background signal which can wash out the low-intensity, high-resolution near-field signal [39].

However, this background results from reflections happening on relatively large length scales compared to the tip apex and thus has lower frequency components than the near-field signal. In other words, we only see significant background contributions at low harmonics of the tip oscillation frequency such as $n = 1$ or $n = 2$. For example, one source of background signal results from the tip shaft oscillating up and down

[3] If it wasn't, then the s-SNOM system would not be providing the desired boost in spatial resolution.

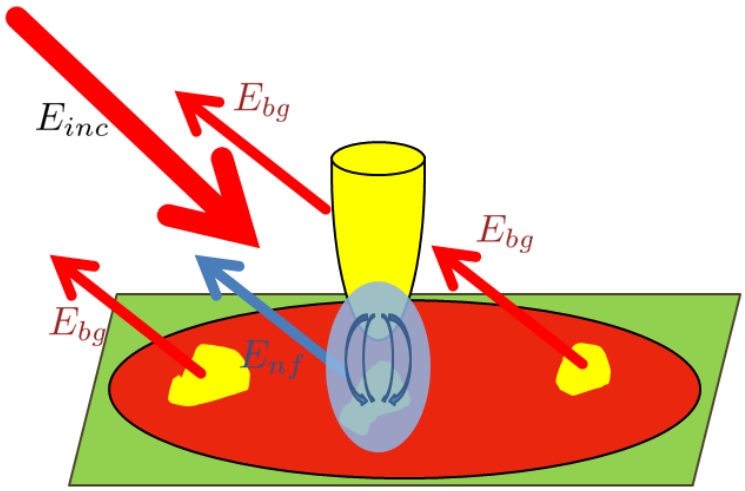


Figure 2.7: This picture shows the contribution of the far-field background. The incident laser light causes and probes a nearfield interaction represented by the blue region. However, the laser spot is much bigger than the near-field interaction region (red circle). This larger laser spot can interact with topographic features on the material surface as well as with the shaft of the tip as well resulting in optical reflections in the same direction as the small near field signal.

through the focus of the laser. This signal is clearly modulated at frequency Ω and will not be detected at higher harmonics. This background elimination component of lock-in detection is its biggest advantage [39]. See Fig. 2.7.

Elastic s-SNOM data is traditionally collected using lock-in detection. This means any given data set corresponds to data resulting from demodulation at a given tip harmonic. Thus, it is customary to speak of e.g. a 3rd harmonic scan, or 2nd harmonic approach curve. In other words, the value of the optical signal demodulated at some tip harmonic **is taken to represent the strength of the near-field interaction**. While this interpretation is valid, it is important to remember what is and is not represented by a demodulated signal. A demodulated signal is the result of the interaction between the non-linear distance dependence of the scattered near-field signal and the oscillatory motion of the probing tip. A demodulated signal does **not** represent the absolute value of the near field signal. In particular, it is not expected that harmonic approach curves will have the same functional form as the raw approach curve. Because of this, it is not necessarily trivial to interpret demodulated s-SNOM data. This is one of the motivations for the real-time acquisition scheme I describe in Chapter 4.

There is another, more significant drawback to lock-in detection. As I have explained, the s-SNOM signal is split into different signals at multiple harmonics. However, in s-SNOM measurements, typically only one or two of these harmonics are measured. This means the information in the other harmonic bands is entirely ignored. This could lead to lower signal-to-noise ratios and lower contrast as well as information about qualitative sample properties being lost.

2.4 Gated Measurement

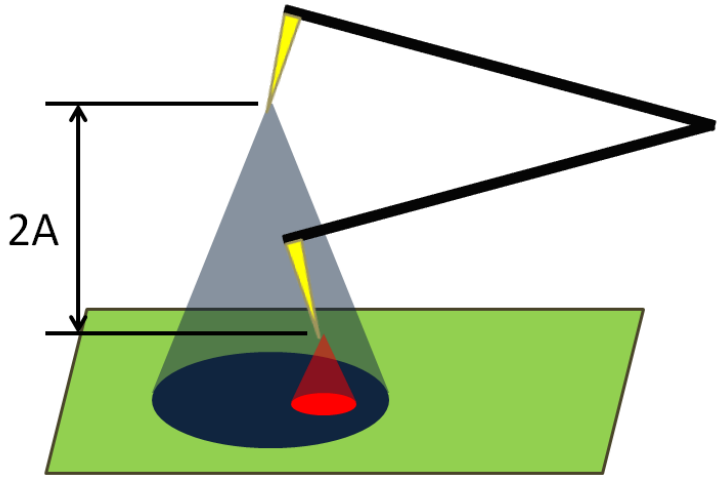


Figure 2.8: In a gated s-SNOM measurement, data is only recorded when the tip is at the top and the bottom of its oscillation. When the tip is closest to the sample, it only interacts with a small region. When it is far from the sample it interacts with a larger region. A is the tip oscillation amplitude.

One alternative measurement technique for s-SNOM is gated measurement. Gating refers to a data flow consisting of temporal bursts of data in between which no relevant information is recorded. In s-SNOM this can be implemented in the illumination or the detection. In gated illumination s-SNOM, the tip apex is illuminated by short pulses of light only during specific portions of its sinusoidal oscillation. In gated detection s-SNOM, the tip is continuously illuminated, but the only signal which is detected corresponds to specific portions of the tip oscillation.

For s-SNOM, gated measurement techniques would involve measuring the signal when the tip is at the top and bottom of its oscillation. The idea is that during a single cycle, when the tip is nearest the surface the scattered signal only contains high resolution information from the local sample near field whereas when

the tip is furthest from the surface, the scattered signal contains information from a larger area of the sample as well as more background. See Fig. 2.8. The optical signal value when the tip is far from the sample is treated as a reference signal and the value when the tip is nearest the sample is treated as the actual signal value. Thus, a value for the near-field interaction strength can be obtained by subtracting these two values from each other. Since data collection is restricted to when the tip is closest to the sample it is thought that this type of technique may lead to higher resolution and better material contrast.

Gated measurements have been successfully implemented in s-SNOM already. In one case, gated illumination was used with a tip-enhanced Raman spectroscopy (TERS) experiment to recreate a distance dependence for the TERS interaction [40]. In another case, a three-dimensional tomographic map of the tip-sample interaction for tip-enhanced fluorescence microscopy was generated by time-stamping the arrival of fluorescence photons at the detector and correlating those with tip position [41]. However, these techniques are inelastic so the detected light is at a different frequency than the incident light. This greatly reduces sources of background as compared to elastic measurements. A gating measurement has not yet been reported for elastic s-SNOM.

2.5 Interferometric Detection

The scattered near field intensity is very small and difficult to detect, especially when it is on top of a large background signal. To boost the near field signal intensity, we utilize interferometric amplification techniques. The idea is that we interfere our near field beam, E_{nf}, with a constant higher intensity amplification beam, E_{amp}, and the resultant cross term is directly proportional to E_{nf} multiplied by E_{amp}, so the near field information is then scaled by the large E_{amp} factor.

$$\begin{aligned} I_{det} \propto E_{tot}E_{tot}^* =& (E_{nf} + E_{amp})(E_{nf}^* + E_{amp}^*) \\ =& E_{nf}^2 + E_{amp}^2 + (E_{nf}^* E_{amp} + E_{nf}E_{amp}^*) \\ =& E_{nf}^2 + E_{amp}^2 + 2|E_{nf}||E_{amp}|\cos(\Delta\phi) \end{aligned} \tag{2.7}$$

Here $\Delta\phi$ represents the phase difference between the near field and the amplification field. The first term in the final line is negligible because its magnitude is very small. The second term has no contribution

at frequencies $n\Omega$ so it is not detected by the lock-in amplifier. Thus, the measured signal comes from the final term which encodes the near field information multiplied by the strength of the amplification field. Note also the dependence on the optical phase difference $\Delta\phi$.

I will now describe three subsequently more advanced interferometric detection schemes. The three schemes vary based on the nature of the amplification field, E_{amp}.

2.5.1 Self-Homodyne Detection

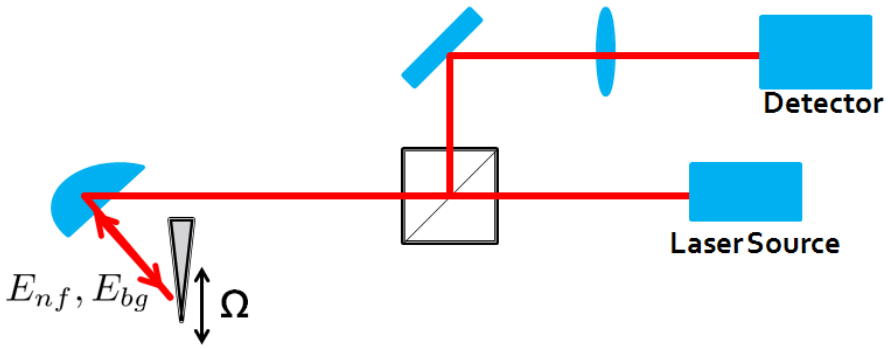


Figure 2.9: In self-homodyne detection, the scattered near field light, E_{nf}, is interfered with the background light scattered from the sample, E_{bg}, to amplify the low intensity near field. Self-homodyne detection provides an image which is a convolution of the near-field amplitude and phase.

In self-homodyne detection, the amplification field comes from the optical reflections from the tip and sample which constitute the background field described earlier. This reflected field is referred to as the far-field background or just background and is called E_{bg} symbolically. In self-homodyne detection, this reflected E_{bg} field takes the place of E_{amp} in the previous discussion.

Self-homodyne detection is the simplest of the interferometric detection schemes and has been used to make many near-field measurements. However, self-homodyne detection has major limitations. The first issue is that E_{bg} essentially arises from (spatially) random scattering events and thus has an uncontrollably varying magnitude depending on where the tip is located with respect to the sample. Furthermore, even if E_{bg} is constant spatially it will not have a constant phase. Thus, because of the $\cos(\Delta\phi)$ factor in the detected signal in Eq. 2.7, if the phase of the near field is constant, then the resultant image will essentially image the phase of E_{bg}. Such an image will exhibit bright and dark interference fringes with spatial periodicity on

the order of λ, the wavelength of the incident light [39].

The fact that both the amplitude and phase of the amplification wave are uncontrollable means that it is possible for many artifacts to manifest in the final image including these interference fringes and other topographical artifacts. Furthermore, the near field has both amplitude and phase. In a self-homodyne measurement only one number (the result of Eqn. 2.7) is measured. Both the amplitude and phase of the near field are convolved together into this one number and, in the self-homodyne setup, it is impossible to deconvolve them.

Subsequent interferometric detection schemes use an amplification beam whose amplitude and phase are known and controlled to overcome the limitations of self-homodyne detection.

2.5.2 Homodyne detection

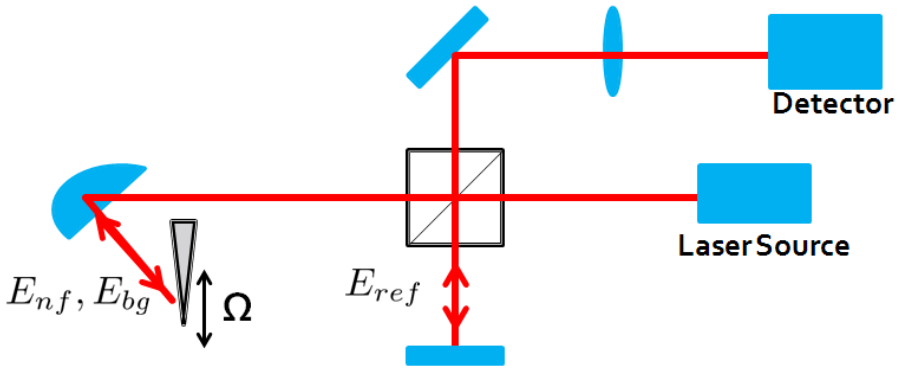


Figure 2.10: In homodyne detection, the scattered near field signal, E_{nf}, is interfered with both the sample scattered background field, E_{bg}, and the constant reference field E_{ref}. if $E_{ref} \gg E_{bg}$ then artifacts resulting from the uncontrollable E_{bg} are suppressed.

In homodyne detection,[4] the amplification beam is a reference beam, E_{ref}, that has been split off from the sample beampath by a beamsplitter and backreflected as in a Michelson interferometer. See Fig. 2.10. There are now three beams interfered at the detector: E_{nf}, E_{bg}, and E_{ref}. E_{ref} should be larger in magnitude than E_{bg} so that the controlled reference beam dominates the uncontrollable background beam.

[4] Unfortunately this name is easy to confuse with self-homodyne. I will be consistent throughout this thesis.

The interference equation reads as such:

$$\begin{aligned} I \propto E_{tot}E_{tot}^* =&(E_{nf} + E_{bg} + E_{ref})(c.c.) \\ =&E_{nf}^2 + E_{bg}^2 + E_{ref}^2 \\ &+ E_{bg}E_{ref}^* + E_{bg}^* E_{ref} \\ &+ E_{nf}E_{bg}^* + E_{nf}^* E_{bg} \\ &+ E_{nf}E_{ref}^* + E_{nf}^* E_{ref} \end{aligned} \tag{2.8}$$

The first term is negligible because it is small in magnitude. The next four terms can be neglected because they do not have a dependence on E_{nf} and are thus not modulated at harmonics of the tip oscillation frequency so they will not contribute to the lock-in detection.[5] The following two terms in the next line are negligible because $E_{ref} \gg E_{bg}$. The remaining terms can be written as such.

$$I \propto 2|E_{nf}||E_{ref}|\cos(\Delta\phi) \tag{2.9}$$

Where $\Delta\phi$ is the relative phase between the near field light and the reference beam.

This is the same equation for the detected intensity as before with E_{ref} replacing E_{bg}. The major advantage, now, is that the magnitude of E_{ref} is constant since it comes directly from the laser source and is simply a beam reflection and $\Delta\phi$ is constant because the length of the beam path does not change in time.

Some reflection reveals that if a measurement is taken at two different values of $\Delta\phi$ it is possible to extract both a value proportional to $|E_{nf}|$ and the value of $\Delta\phi$. Consider two measurements taken at some unknown $\Delta\phi$ and $\Delta\phi - \frac{\pi}{2}$.

$$\begin{aligned} I_1 \propto& 2|E_{nf}||E_{ref}|\cos(\Delta\phi) \\ I_2 \propto& 2|E_{nf}||E_{ref}|\cos(\Delta\phi - \frac{\pi}{2}) = 2|E_{nf}||E_{ref}|\sin(\Delta\phi) \\ I_1^2 + I_2^2 \propto& |E_{nf}|^2 \\ \arctan\left(\frac{I_2}{I_1}\right) =& \Delta\phi \end{aligned} \tag{2.10}$$

Thus, by making two measurements at a single point and performing the preceding calculations, it is possible to determine a value for the magnitude and phase of the near field at a given point. These values

[5] This is, in fact, an approximation. In reality E_{bg} does have some modulation at frequency Ω and higher harmonics. However, its effect decreases substantially with increasing harmonic number, n.

can be compared to values at other points in an image to see spatial near-field optical contrast in both phase and magnitude.

The addition of $\frac{\pi}{2}$ to the optical phase difference can be accomplished by using a piezo to adjust the length of the reference arm by a small amount. A phase change of $\frac{\pi}{2}$ corresponds to a beam path difference of $\frac{\lambda}{4}$. Since the beam travels back along its same path this means the mirror must be shifted a distance $\frac{\lambda}{8}$ which is on the order of hundreds of nanometers or microns depending on the laser source.

The homodyne detection scheme is very powerful because it provides a additional information and advantages over self homodyne, yet is still relatively simple. Homodyne arguably provides the best optical contrast and signal-to-noise ratio of the interferometric detection schemes.

2.5.3 Pseudo-heterodyne detection

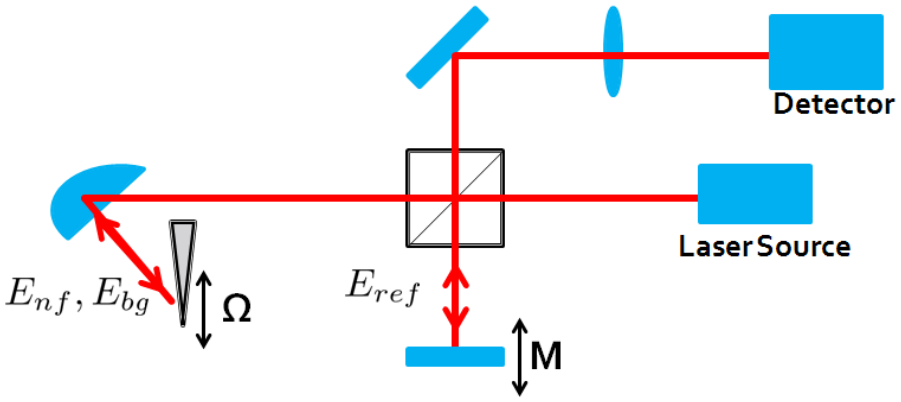


Figure 2.11: In pseudo-heterodyne detection, the reference mirror is oscillated at frequency M. E_{nf} is modulated at frequencies $n\Omega$. The result is that these two signals are mixed at the detector and the signal splits into sidebands located at frequencies $n\Omega + mM$. The sidebands contain amplitude and phase information and, if chosen correctly, do not contain undesirable background signal.

Pseudo-heterodyne[6] detection utilizes a reference arm, like in homodyne detection, to amplify the weak near field signal. However, in Pseudo-heterodyne detection the optical phase of the reference arm compared to the phase of the signal arm, $\Delta\phi$, is sinusoidally modulated. This is accomplished by applying a sinusoidal signal to the piezo on which the reference mirror is mounted. The reference field now takes on

[6] This detection scheme is, again, somewhat poorly named. The name pseudo-heterodyne is almost purely historical.

this form.

$$E_{ref} = |E_{ref}|e^{i\omega t + i\gamma\cos(Mt)} + c.c. \tag{2.11}$$

M is the frequency of the reference mirror oscillation and is typically on the order of hundreds of hertz. γ is the modulation depth, or amplitude, of the reference mirror oscillation.

A series of calculations found in Ref. [42] show that when this reference beam is interfered with E_{nf} –which has Fourier components at frequencies $n\Omega$– the result in Fourier space is that each peak at frequency $n\Omega$ is split into sidebands separated by frequency M. Intuitively, since the detector measures intensity which is proportional to E^2, it mixes the two signals at frequencies Ω and M. In other words we detect signal at frequencies $n\Omega + mM$, where n and m are integers (m can be negative).

If γ is carefully chosen, then it is possible to determine a number proportional to $|E_{nf}|$ and calculate $\Delta\phi$ by recording the amplitude of two of these sidebands in Fourier space and performing calculations similar to those in the homodyne case on these two values. In practice we record sidebands with values $m = 1$ and $m = 2$.

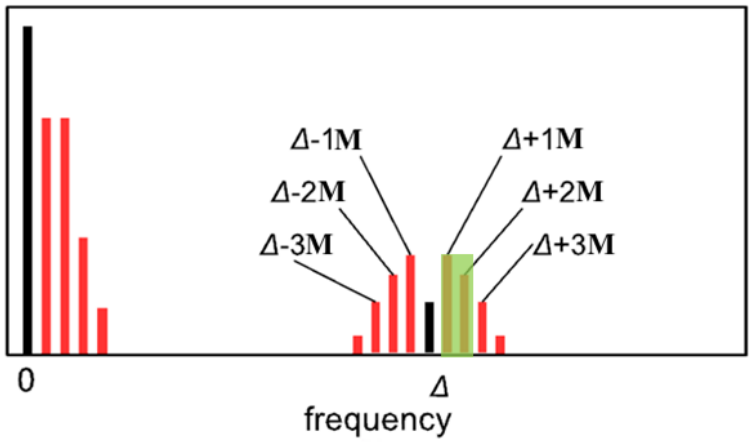


Figure 2.12: In pseudo-heterodyne detection, the oscillation of the reference mirror at frequency M causes the harmonic frequency signals to split into sidebands separated by frequency M. Monitoring the indicated sidebands provides information about the amplitude and the phase of the near-field signal. Figure adapted from Ref. [34].

Pseudo-heterodyne is an interesting and powerful measurement technique. Its main advantage over homodyne detection is that it is able to measure the amplitude and phase of the near field in a single scan as opposed to homodyne which requires two scans for this information. It is possible to interpret the oscillation

of the reference mirror as performing a continuously varying homodyne measurement.

However, pseudo-heterodyne has drawbacks as well. It is a somewhat complicated measurement technique and thus has lots of opportunities for sources of noise. One prime candidate for noise is mirror instability or drift. Pseudo-heterodyne also requires an expensive lock-in amplifier which can simultaneously follow two reference signals and preferably perform real time math on its input signals.

However, the biggest disadvantage of pseudo-heterodyne is that, even more so than general lock-in s-SNOM, it neglects a large portion of the actual signal falling on the detector. This is because each sideband of each tip harmonic peak carries information about the scattered near field signal, yet in pseudo-heterodyne detection we only monitor two of these sidebands. The result is that the images captured by pseudo-heterodyne detection have less contrast and higher signal-to-noise ratios than, for example, homodyne images.

2.6 Summary

In this section I have given a theoretical presentation of all of the aspects of s-SNOM that allow it to provide sub-diffraciton limited material contrast. In particular I have presented three theoretical models for the tip-sample interaction and three interferometric detection schemes. The goal of this thesis is to investigate an alternative detection system based on real-time data acquisition of the optical signal as opposed to lock-in detection. Part of the motivation for this new detection technique is the criticism that lock-in detection rejects large percentages of the signal intensity, whereas a real-time acquisition scheme records all signal information. I also hope to use data captured with my real time acquisition technique to investigate the validity of the tip-sample interaction models I have presented.

Chapter 3

Experimental

3.1 AFM

Combining spectroscopy with SPM, any s-SNOM setup is designed around an AFM or STM which serves as its foundation. Effective design and use of s-SNOM systems relies on an understanding and mastery of that instrument. Different groups have used both commercial and home-built SPMs as the foundation for their s-SNOM systems. My s-SNOM setup is built around a commercial AFM from Anasys Instruments.

This AFM can be operated in tapping or contact mode, however, I operate exclusively in tapping mode since traditional s-SNOM detection techniques rely on locking in to the tip oscillation frequency. This AFM offers an overhead CCD camera which provides an overhead view of the cantilever and sample to aid in sample alignment. It has also been designed so that the tip remains stationary while the sample is scanned in the x, y, and z directions to create topographic images. The reason for this is so that the external s-SNOM laser can be focused once onto the tip apex and does not need to be realigned during scanning.

To maintain constant tip-sample separation, the AFM uses a feedback laser to monitor the tip oscillation amplitude. This laser is shined onto the back of the cantilever and deflected onto the center of a two section photodiode. As the cantilever bends up and down, the laser spot is deflected into one section of the photodiode or the other. The difference between the laser intensity on the two photodiode section is monitored. This signal is known as the tapping signal. The amplitude of this oscillation is monitored by a feedback lock-in amplifier in the AFM which locks in to the tip oscillation frequency Ω. This signal is known as the feedback lock-in amplitude. The feedback lock-in amplitude is an oscillatory signal whose amplitude

is proportional to the tapping amplitude of the cantilever. This is because during a larger tip oscillation the cantilever bends more which results in a bigger deflection of the feedback laser which results in a larger difference between the photodiode sections.

To control the tip-sample separation, the feedback lock-in amplitude is compared to a software or user defined set point. Far away from the sample, the tapping amplitude is constant and largest. As the sample is brought closer to the tip, the tapping amplitude (and correspondingly the feedback lock-in amplitude) decreases due to Van-der-Waals interactions between molecules in the tip and sample. See Fig. 3.1. The value of the feedback lock-in amplitude at the desired tip-sample separation distance is chosen as the value for the set point. Force approach curves depicting feedback lock-in amplitude versus tip-sample distance can be used to calibrate the feedback lock-in amplitude voltage to an actual tip oscillation amplitude in nanometers.

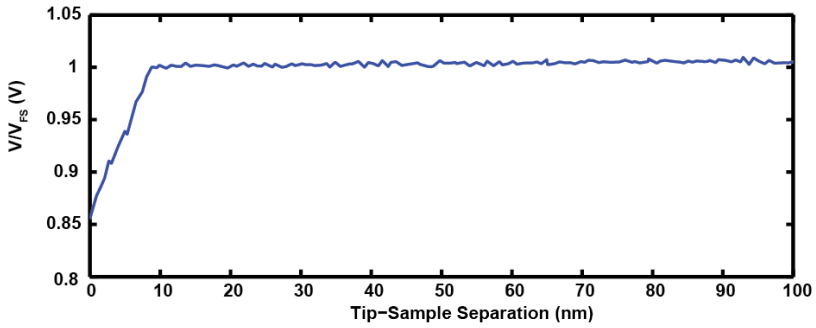


Figure 3.1: At large tip-sample separations there is no interaction between the tip and the sample so the oscillation remains constant. However, when the tip gets close to the sample and begins interacting the tapping amplitude begins to decrease. The set point is chosen to correspond to a small tip-sample separation for which the tip and sample are interacting. The PID closed feedback loop adjusts the tip-sample separation in real-time to ensure the feedback lock-in amplitude remains constant. V_{fs} is the free space tapping amplitude far from the sample. V is the amplitude of the tapping signal.

To achieve stable feedback, the AFM uses a z-piezo to bring the sample towards the tip while a PID feedback loop monitors and attempts to minimize the difference between the feedback lock-in amplitude and the set point. Once AFM is in stable contact, the sample can be scanned in the x and y directions with different piezos and the PID loop will respond to differences in sample height by applying voltages to the z-piezo to ensure the feedback lock-in amplitude equals the set point. Monitoring the z-piezo voltage and correlating it with the x and y piezo voltages provides a 2D image of the sample topography.

3.2 Parabolic Mirror

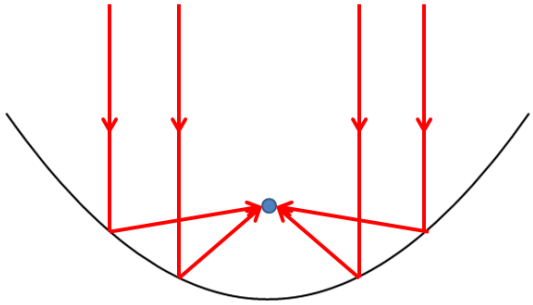


Figure 3.2: A parabolic mirror focuses all beams parallel to the focal axis of the mirror towards the focus of the parabola.

In s-SNOM, a laser (distinct from the AFM feedback laser just described) is focused into the region containing the AFM tip apex and the sample. Light scattered from this region must be recollected and sent to a detector. It has become customary to use one parabolic mirror for both focusing and collection of light. The major reason for this is that parabolic mirrors offer large numerical apertures which captures a large portion of the already small near-field scattered signal [43]. The large numerical aperture also allows focusing of the incoming beam into as tight a diffraction-limited spot as possible which is advantageous for various reasons. Parabolic mirrors are also nice for alignment because if the incident beam is parallel to the focal axis of the parabolic mirror, it will always be focused to the focus of the parabola. See Fig. 3.2.

My setup utilizes an approximately 0.4 numerical aperture off-axis parabolic mirror from Newport. The parabolic mirror is mounted onto a micrometer XYZ translation stage so that the laser can easily be focused onto the tip apex with the help of the overhead CCD camera on the AFM. The parabolic mirror is aligned to the laser beam path by replacing the parabolic mirror with a flat mirror whose reflective surface is perpendicular to the parabolic mirror's focal axis. Then the tilt of this mirror is adjusted to ensure the reflected beam is colinear with the incident beam. This ensures the beam is parallel to the parabolic mirror's focal axis and thus that the parabolic mirror is properly aligned.

3.3 Laser Sources

For my s-SNOM illumination I utilized a $10.8\mu m$ carbon-13 $^{13}CO_2$ gas laser. The output power is 750 mW and was attenuated to approximately 10 mW at the tip. This laser operates in the infrared and is thus invisible to the human eye. To align one this invisible beam I made it colinear with a visible Helium Neon ($HeNe$) laser ($\lambda = 632nm$) and then align the $HeNe$ with the assumption that if the $HeNe$ is aligned then the CO_2 laser is also aligned.

3.4 Detector

My spectroscopy experiments are all entirely elastic meaning the only frequency of light involved in the measurement is that of the laser source, so it suffices to detect power at a single frequency. To detect the infrared CO_2 laser light, I utilize a Judson $HeCdTe$ (mercury cadmium telluride or MCT) photo-detector. An MCT detector is a semiconducting photo detector device that converts incident light intensity into an electrical current. The MCT has a 250 μm active region and a cutoff frequency at 14 μm. It is mounted on an XYZ micrometer stage for alignment. Another parabolic mirror is used to focus the incident beam onto the MCT active area.

3.5 Signal Processing

3.5.1 Lock-In Amplifier

In s-SNOM, the optical signal is modulated at the tip oscillation frequency Ω and integer multiples $n\Omega$ thereof. Because of this it makes sense to only monitor these tip harmonics. The most effective way to monitor a signal at a set frequency is to use a lock-in detector.

A lock-in amplifier compares one AC electronic signal to a reference signal. The amplifier uses an electronic circuit called a phase locked loop to calculate the Fourier component of the signal of interest at the frequency of the reference signal.

In the case of s-SNOM, the reference signal is the tapping signal from the AFM at frequency Ω and the signal of interest is the optical signal from the MCT. The lock-in amplifier used in my setup is a Zurich

lock-in. It has two inputs that can be used as reference signals. It can also lock in on integer multiples of these input signals as well as add the reference frequencies.

In my setup, the lock-in processes the signal and calculates a value to feed back into the AFM software so that the near field signal value can be correlated with a physical spot on the sample by the AFM software.

3.5.2 Data Acquisition Card

The primary goal of this thesis is to investigate the use of a data acquisition card (DAQ) to record the real time s-SNOM signal as opposed to a lock-in detector. One disadvantage of lock-in detection is that it only monitors one frequency component of the signal and ignores all of the rest. This leads to lower signal-to-noise ratios than may be possible. Furthermore, the interpretation of lock-in data is not as transparent as real-time signal.

To record the real time s-SNOM signal output from the MCT detector, I used a National Instruments (NI) DAQ. The DAQ has four acquisition channels. Each channel can record data at 10 MHz and can buffer up to ten million datapoints. In my setup, depending on the experiment I am running, I acquire from various combinations of the optical signal from the MCT, the tapping signal, the z-piezo position, and the x-scanner position. I use LabView code to control the DAQ and to extract the raw data off of the DAQ. The raw data consists of multiple large data files which I ultimately process in MatLab.

The major advantage of this type of data processing is that it analyzes all of the available data and, thus, loses no information. As a result, this type of data acquisition is extraordinarily valuable as a diagnostic technique to assess the performance of my s-SNOM setup. One drawback is that the data processing cannot be performed in real time and it is not yet possible to create 2d images with the DAQ signal processing.

3.6 Interference Arm

Different s-SNOM modalities require that the scattered light from the sample is interfered with a strong reference beam of known optical phase. To accomplish this, the beam is passed through a beamsplitter and one arm is sent to the tip and sample and the other arm is sent to a standard flat mirror that reflects the beam directly back along its original path. This reflected beam and the signal beam coming from the scattered

near-field light then pass through the beam splitter again but are this time recombined and interfere with each other before passing into the MCT detector. See Fig. 2.11.

In pseudo-heterodyne detection scheme, the signal beam is interfered with a beam whose phase is varying sinusoidally. To achieve such a reference beam, the reference mirror is mounted onto a piezo whose length can be controlled by an external sinusoidal signal. The length of the reference arm must be the same length of the signal arm to higher accuracy than the coherence length of the source laser to ensure the two interfered beams remain coherent. In pulsed laser measurement the two arms must have the same length to higher accuracy than the length of the laser pulse to ensure both pulses fall on the detector simultaneously.

3.7 Operation

Operating a s-SNOM instrument essentially requires focusing a laser beam on the apex of an AFM tip using a parabolic mirror and operating the AFM. In this section I will present a brief account of this process in more detail. Though I rebuilt this setup from the table up, I will only explain how to take near-field measurements once all preliminary alignments have been performed and only tune-up alignments need to be performed.

Since the CO_2 laser used for the s-SNOM measurements is infrared, it is impossible to align its beams by eye since the human eye cannot detect infrared light. To overcome this, the first step in alignment is to make the infrared laser colinear with the $HeNe$ laser. Once the two beams are colinear it is possible to align the $HeNe$ laser by eye and safely assume that the infrared laser follows the same beam path. The two beams are made colinear by first passing the infrared beam through a beam combiner (transmissive in IR, reflective in visible) and reflecting the $HeNe$ off of the same beam combiner. The two beams are then deflected into a long auxiliary arm and independently passed through two irises to guarantee they follow the same beam path. The $HeNe$ is aligned through the irises by eye and the CO_2 laser is aligned by placing a power meter behind each iris and maximizing the power. If two beams pass through two common points they are necessarily colinear.

Next I must ensure that reference mirror reflects light directly back along the original beam path. This is done by checking that the backreflection passes through the center of an iris that the original beam

passed through. Next I align the beam to the focal axis of the parabolic mirror by a similar flat mirror backreflection technique described in sections 3.2.

The laser must then be focused onto the tip apex by translating the parabolic mirror in the x, y, and z directions. The advantage of the parabolic mirror is that if the input beam is parallel to the focal axis then it will focus properly no matter the relative XYZ position between the beam and the mirror. The focus is brought to the tip by observing the $HeNe$ laser spot falling onto a sample near the AFM tip (at this time the tip is microns away from the sample, i.e. not in contact). The mirror is translated until the focus falls on the tip.

Now the AFM tip is brought into contact with the sample surface by raising the sample until the AFM feedback indicates contact has been made.

Once the tip is in contact, if it is over a sample with a large near-field signal such as gold, it is likely that I can begin to measure signal with the MCT and lock-in. The scattered light from the tip-sample interaction is collected by the parabolic mirror and directed into the detection arm where it is focused onto the MCT. The MCT then sends the signal to the lock-in detector which measures the components of the signal at the various tip harmonics. The value of these harmonics is related to the strength of my signal.

At this stage, I translate the XYZ position of the MCT to maximize these signals to ensure the MCT is aligned well. Next, I adjust the XYZ position of the parabolic mirror to fine tune the alignment of the laser to the tip apex to maximize the near-field signal. Typically I maximize the n=3 signal since it is inadvisable to maximize low harmonics since that may actually maximize a source of background. Sometimes I can see signal at up to the $n = 10$ harmonic.

Next, I setup the system for whichever detection scheme I am choosing to use. This requires blocking or unblocking the reference mirror and either leaving it stationary or oscillating it with the piezo. I also must do various rewirings to ensure either that the lock-in is detecting the proper input signals and reference signals or that the DAQ is recording the proper channels.

Once all of this is setup, I am ready to begin performing scans with the AFM. The data is collected on the computer through various combinations of the lock-in software, AFM software, and LabView software.

I can perform a 2D XY scan which give an image of the sample surface, I can perform 1D line scan

which characterizes sample changes in a given direction, or I can perform 1D approach curves which take measurements at a single point above the sample as the tip-sample distance is varied. After the data is recorded, I can use various data analysis programs such as Gwyddion, Matlab, or Mathematica to analyze the data and extract the relevant information.

Chapter 4

Results

The goal of my work was to develop a new detection system for s-SNOM that would provide new information about the tip-sample interaction and lead to improvements in both the technique and theory behind s-SNOM. I chose to do this by developing a detection scheme which records the optical signal in real time as opposed to recording a time-averaged demodulated signal.

This real-time detection scheme is motivated by various ideas. The first motivation for real-time detection is that it simply provides more information about the signal. In fact, real-time detection provides all of the possible information that could be collected by the MCT detector. As a result, any data which would be collected by other means such as lock in or gated detected can be simulated from real-time data. This immediately makes it a versatile and inclusive measurement technique. Real-time data may also contain information that cannot be obtained by other means. Furthermore, it is also possible to collect other signals related to the tip position simultaneously with the optical signal. These signals can then be correlated to each other. Direct correlation of tip position and optical signal represents a major leap in s-SNOM technology.

However, this scheme was motivated by more than the abstract idea that it provides more data. I intended to use this new scheme to investigate new methods which may provide higher resolution and better material contrast. I wanted to see if the additional data somehow contained some high resolution information which could be extracted. To this end, I performed an optical approach curve and used the data to simulate gated detection to determine if gated detection will provide high resolution information as predicted.

The additional information provided by real-time detection over lock-in detection also provides another lucrative potentiality. Since the data from real-time detection is so complete as well as simple, it is very

straightforward to analyze and interpret. This makes it a very useful tool to analyze the performance of the s-SNOM setup in which it is implemented. It is possible to directly analyze theories of s-SNOM operation. To demonstrate this I recreate, for the first time, a raw optical approach curve for elastic s-SNOM from my real-time data. This is as opposed to harmonic demodulated optical approach curves whose interpretations are somewhat opaque. I use this raw optical approach curve to determine the validity of one of the models of the tip-sample interaction presented earlier.

It should be noted that much of my analysis relies on simulating lock-in data from my real-time data. My techniques could be criticized for providing the same information as lock-in data. However, some of my analysis would be possible but extremely difficult with a lock-in detector.[1] Furthermore, even if some real-time data could just as easily be taken from lock-in detection it is important to note two disadvantages of lock-in detection. First, the Zurich lock-ins used in the Raschke group can only record up to 6 harmonics of the reference signal, whereas the number of harmonics measurable by lock-in detection is only limited by the 10MHz sampling rate of the DAQ. Second, the Zurich lock-ins are highly advanced instruments costing approximately $20,000 whereas the DAQ is fairly simple and costs approximately $4,000.

There are two major criticisms of the real-time detection technique as it is currently implemented. The first criticism is that it is only possible to record one dimensional sample images. This is related to the number of acquisition channels on the DAQ as well as time of measurement and the difficulty of correlating, in time, a large number of variables. The second criticism is that the real-time data requires a large amount of post-processing whereas the lock-in techniques immediately feed data into the AFM software which automatically correlates all of the different variables.

4.1 Real-Time Acquisition Scheme

As described ealier, I developed a real time detection scheme for s-SNOM based on a NI DAQ. The DAQ replaces the lock-in amplifier from traditional s-SNOM measurements. It has four acquisition channels that can acquire ten million pixels of data at 10MHz.

I have the capability to monitor the optical signal from the MCT, the voltage that is applied to the

[1] For example the raw approach curve reconstruction. See section 4.5.1.

z-piezo, the voltage that is applied to the x-piezo and the tapping signal. The combination of the z-piezo and the tapping signal tell me the absolute tip-sample separation. The x-piezo tells me the x position of the tip relative to the sample during a line scan. The MCT signal tells me the optical signal at a given time.

The DAQ is controlled with LabView software. The software communicates with the DAQ to tell it to begin acquiring and ,when the acquisition is complete, the DAQ dumps the data into LabView. The raw data consists of one data set for each acquisition channel with up to 10 million points separated by time steps as small as 10ns. LabView can save this data into standard data formats which can then be processed by MatLab or Mathematica.

All of the measurements reported in this thesis were recorded using self-homodyne interferometric detection. This is primarily because self-homodyne is the simplest and most reliable scheme. It is possible to make use of any of the interferometric schemes depending on the purpose of the experiment.

4.2 Measurement Technique and Raw Data

The main measurement taken for this thesis is the approach curve measurement. An approach curve is taken by removing the tip a set distance from the sample and then slowly moving the sample towards the tip until the tip comes into contact with the sample.

During this process I use the DAQ to simultaneously monitor the real-time signal from the MCT, the z-piezo voltage and tip tapping amplitude. The MCT signal gives me a real-time near-field signal and the other two signals give me information about the exact tip-sample separation.

The main approach curve data set shown in Fig. 4.1 was taken with a platinum coated silicon tip with a resonant frequency of roughly 250 kHz. The tapping amplitude was approximately 100nm. The distance of the approach curve was 600nm.The DAQ was set to record ten million pixels at 5 MHz on each acquisition channel. The tip approached a flat gold sample.

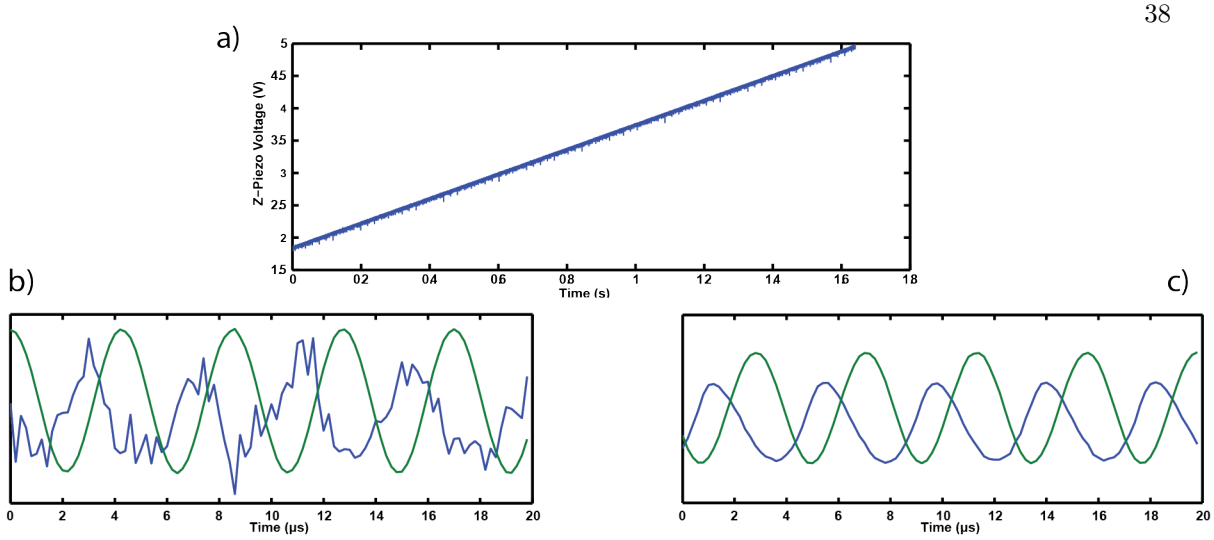


Figure 4.1: Raw data from real-time acquisition detection. a) z-piezo voltage during the course of an approach curve. The linearly increasing voltage represents the sample moving at a constant velocity towards the tip. b),c) tapping (green) and optical (blue) signals for a few cycles of the tip oscillation. In b) the tip is far from the sample and we only see a weak, noisy background optical signal. In c) the tip is in contact with the sample and the optical signal is much stronger and cleaner. The optical signal is slightly non-linear as expected.

4.3 Lock-In Data Simulation

I will first describe how I simulate lock-in harmonic detection from my data. The lock-in can be thought of as collecting signal for an amount of time specified by its user-defined time constant and then Fourier transforming that data set and measuring the value at a frequency specified by the reference signal. This calculation is performed quickly by the use of a fast Fourier transform and can be correlated with the z-piezo voltage –which is monitored by the AFM in real time– thus assigning a value for the harmonic demodulated optical signal to each value for z, the tip-sample separation. This is how demodulated optical approach curves have traditionally been measured, but note that it is not a direct recreation of the distance dependence of the near-field scattering, but rather a composition of this distance dependence with the tip-oscillation and harmonic demodulation. This is the process I simulate with my data.

To simulate harmonic approach curves, I bin my data similarly to how the lock in time-bins the input signal by its time constant. I bin each data set into bins a certain number of pixels long. This corresponds to collecting data for a certain amount of time.[2] Now I must assign to each bin a value for the tip-

[2] The number of a pixels in a bin can be directly translated into a respective time constant for lock-in detection. The lock-in

sample separation and a value for the optical signal measured at any tip harmonic, $n\Omega$. For the tip-sample separation, I look at the z-piezo voltage for a given bin, take the average value, convert the voltage to a distance and assign that distance to that bin.[3]

To assign values for the optical signal to each bin, I need to know the tip oscillation frequency. I calculate this frequency, Ω, by finding the peak value of the Fourier transform of the tapping signal since the tapping signal is a clean sinusoidal signal at frequency Ω. Then, I take the real-time optical signal in a given bin and take its Fourier transform. I then record the value of this Fourier transform at values $n\Omega$ up to the desired harmonic and correlate those values with the given bin.

I have now correlated to each bin a tip-sample separation and a value for the optical signal demodulated at a range of tip harmonics. Plotting these two values against each other results in Fig. 4.2. This figure simultaneously shows multiple optical approach curves for various tip harmonics. These simulated harmonic approach curves agree with what would be expected for harmonic approach curves directly measured with lock-in detection since in both cases the data is ultimately analyzed in the same way. In particular, notice that the higher tip harmonics have a shorter distance dependence. Harmonics with a shorter distance dependence indicate interactions on shorter wavelengths and thus encode higher spatial resolution information. Even though I am simply simulating data that could be taken with lock-in detection, my detection technique still has the advantage that I am able to measure all harmonic signals simultaneously in a single shot.

I will compare the gated illumination simulation to these approach curves to determine if it has a steeper distance dependence and thus if it provides higher spatial resolution than the lock-in method.

time constant is chosen so that it is shorter than the pixel dwell time for the AFM.

[3] I set 0 nm separation to be the point when the z-piezo voltage is highest for the approach curve. The conversion between V and nm was calibrated earlier.

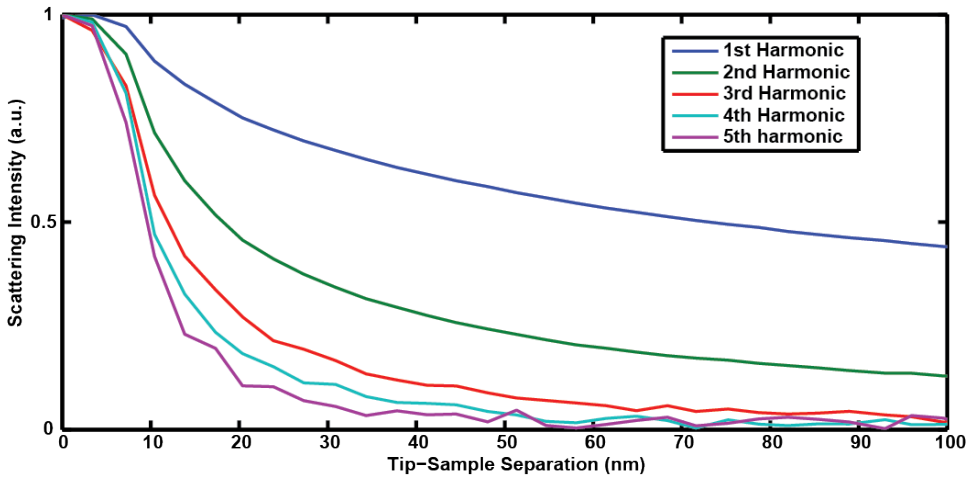


Figure 4.2: Five normalized simulated harmonic approach curves. Notice that the higher harmonic approach curves have a steeper distance dependence. Also notice the tail for the lower harmonics approaches some non-zero value as the tip-sample separation increases. This is related to the background signal.

4.4 Gated Detection Simulation

A gated detection s-SNOM measurement would require recording optical data for the tip at the bottom and top of its oscillation cycle for a specified amount of time. Next, the signals levels corresponding to adjacent 'top' signals and 'bottom' signals would need to be subtracted. This difference is the optical data point for this type of measurement. It is also possible to calculate this difference and average it over many cycles. To create a gated optical approach curve, these optical data points would need to be correlated with the tip sample separation.

Just as in the case for lock-in simulation, I bin the data into time bins and assign a tip-sample separation to each bin in the same way. However, I must calculate the optical signal differently. The tapping signal dataset contains information about the tip's position in its oscillation cycle. The idea I use here is to find pixels for when the tip is at the top and bottom of its cycle by finding maxima and minima of the tapping amplitude signal and then recording the optical value at the corresponding pixels and making the necessary calculations.[4]

I can perform this process for all (or many) of the tip oscillation cycles in a given bin and average

[4] Due to electronic delays in the signal coming from the MCT and the AFM feedback photodiode, there is a phase offset between the tapping and the optical signals visible in Fig. 4.1 b) and c). I wrote MatLab code to determine the value of this offset and subtract it away.

over all of those values. This gives me a value for the optical signal for a given bin that can be correlated with the tip-sample separation for that bin to produce a gated detection optical approach curve.

4.4.1 Results and Discussion

In Fig. 4.3 a), I plot the simulated gated optical approach curve versus the simulated lock-in optical approach curves for tip harmonics $n = 1, 2, 3$ (all normalized). It is clear that the gated illumination curve has the exact same behavior as the first harmonic curve. This shows that the gated illumination method provides no more information than the lock-in detected 1st harmonic. This is unfortunate because the 1st harmonic signal carries a very large background component that we wish to eliminate from our measurements.

This could have been predicted by interpreting the processes involved in the two different measurement techniques. Fig. 4.4 depicts the tip oscillating in the z direction and how this is equivalent to sweeping left and right along the near-field scattering distance dependence curve. The gated detection method records the values at the top and bottom extrema of this sweep and then takes the difference. This is exactly analogous to measuring the secant curve to the near-field scattering distance dependence curve, or taking the derivative of the curve in the limit that tip oscillation amplitude goes to 0. Similarly, the process of taking the Fourier transform component at frequency 1Ω is the same as taking this sweep and weighting it heavily positively at the top of the sweep and heavily negatively at the bottom of the sweep.

Thus, both the gated detection technique and lock-in detection at the 1st harmonic amount to measuring the 1st derivative of the near-field scattering distance dependence. This has proven to be a valuable insight in the interpretation of lock-in harmonic demodulation data and was the motivation for my near-field scattering distance dependence reconstruction technique. It is non-trivial (if at all possible) to extend this derivative intuition to higher harmonic demodulations of the optical signal. Another possible way to see the equivalence of the two techniques is through the Nyquist sampling theorem which implies that sampling a signal twice per oscillation cycle only gives information about the first harmonic of that signal.

The analysis of this data provided the valuable knowledge that this gated detection technique provides no more information than a 1st harmonic lock-in detection scheme. However, it is possible that a gated detection scheme is still worth pursuing. It may be the case that we can extract near-field information by

using a gated illumination technique that is more complex than simply measuring the signal at the top and bottom of the tip oscillation cycle. Furthermore, the pulsed nature of the measurement invokes ideas for non-linear and ultra-fast s-SNOM experiments.

In this particular experiment the tapping amplitude was very large compared to the expected near-field length scale. It is possible that this is why the longer-distance-varying far-field background is so predominant in the signal. In Fig. 4.3 b), I analyzed the real-time data to simulate gated detection for a much smaller tapping amplitude which should provide higher spatial resolution. Instead of taking the signal when the tip is at the top and bottom of its oscillation, I took the signal when the tip was at the bottom of its oscillation and a point when the tip had just barely moved away from the bottom of its oscillation and subtracted the optical signal for these two points. This is the equivalent of doing gated detection with a smaller tapping amplitude. It is clear that this makes the distance dependence of the gated detection curve steeper. This is because a smaller tapping amplitude probes the quickly decaying near-field rather than the slowly varying far-field background. However, using a smaller tapping amplitude would also steepen the distance dependence of the 1st harmonic signal and I predict that even for the case of a smaller tapping amplitude the gated detection would agree with the 1st harmonic lock-in data for the reasons explained earlier.

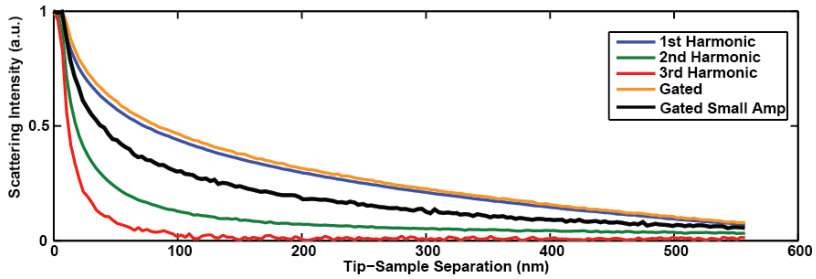


Figure 4.3: Plot of the simulated gating approach curve compared to simulated harmonic demodulated approach curves. Notice the very close agreement between the gating approach curve and the 1st harmonic approach curve. The black line shows the small amplitude gated detection. Notice that it has a steeper distance dependence than the first harmonic but it is not steeper than the second harmonic for the original tapping amplitude lock-in detection.

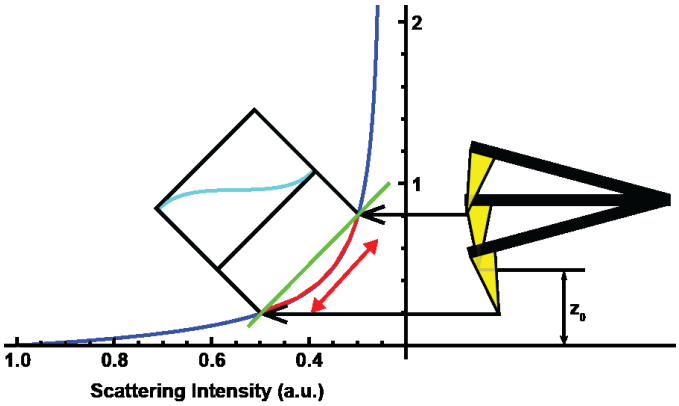


Figure 4.4: This figure depicts how a gated measurement and 1st harmonic measurement record the 1st derivative of the scattering intensity. At a constant tip-sample separation, z_0, the tip sweeps back and forth through the red region. A gated measurement only measures the values at the extrema and thus measures the slope of the green secant line. In a 1st Harmonic measurement, when the signal is Fourier transformed the red region of the Scattering curve is weighted by the displayed sinusoidal curve and integrated. One extrema is weighted positively and one negatively so the integration results, approximately, in calculating the difference between the two extrema.

4.5 Near-field Scattering Distance Dependence Reconstruction

A thorough understanding of the tip-sample interaction is critical to the continued development of s-SNOM as a measurement technique. This is because it is through this scattering mechanism that material contrast is determined so the better this interaction is understood the more properly s-SNOM data can be interpreted. The key to understanding the tip-sample interaction is through optical approach curves since they contain the relevant information of the interaction.

I have presented two theoretical models that predict this near-field signal distance dependence: the simple dipole model and the extended dipole model. I explained the merits and drawbacks of each model in sections 2.2.2 and 2.2.3. However, like all theories these are only approximations. Furthermore, the fundamental prediction of these models –the distance dependence of the near-field scattering– has never been directly measured.

It is possible to make indirect measurements of the models. For example, it is possible to use an approach curve based on one of these theoretical models and simulate the effect of harmonic demodulation of this curve as an oscillating tip sweeps through it. This signal can then be compared to experimental lock-in detected measurements [38]. However, this sort of measurement is in some sense one degree of separation away from a measurement of the actual near-field scattering distance dependence. In this section I present a new and surprisingly simple technique to recreate the near-field scattering distance dependence from real-time s-SNOM data.

4.5.1 Data and Analysis

I use the same data set as that for the gated illumination approach curve. In particular, the necessary detection channels on the DAQ are the z-piezo voltage, the tapping signal, and the optical signal from the MCT. Self-homodyne interferometric detection is utilized for this measurement.

As explained in section 2.3, the signal from the MCT is best interpreted as $I(z(t))$ where $I(z)$ represents the near-field scattering distance dependence function and $z(t)$ represents the tip sample separation as a function of time. With the real-time DAQ detection system, I am able to measure $I(t) = I(z(t))$ and $z(t)$.

To back out $I(z)$ from these two measurements I parametrically plot $I(t)$ versus $z(t)$. This is the simple idea behind my approach curve reconstruction technique.

In the previous sections it was sufficient to use only the z-piezo voltage to represent the tip sample separation for a given time bin. This is because the lock-in detection and gated illumination detection average over the optical approach curve as the tip oscillates around a given central tip sample separation z_0. The result of this averaging can then be correlated to this specific tip sample separation z_0. This averaging is precisely the process I want to avoid since it eliminates (or at best conceals) near-field scattering distance dependence information.

I avoid averaging my signal in this fashion by using the z-piezo voltage and the tapping amplitude signal to trace the real-time tip position. At each pixel I convert the z-piezo voltage to a center point z_0 of the tip oscillation and add to it the tapping amplitude value at that point scaled by the measured tip oscillation amplitude. This correlates an exact tip-sample separation to each pixel. Now I must correlate to each pixel a value for the optical signal.

Initially it may seem that it would be possible to directly assign the optical signal from the MCT to each pixel, however there is a subtlety which prevents this straightforward identification. This issue has thwarted attempts to reconstruct raw optical approach curves so it is worth discussing.

The subtlety arises from the fact that the MCT detector is AC-coupled. This means that only high frequency components of the optical signal falling on the MCT are converted into an electrical signal. DC components are subtracted off. Fig. 4.5 provides a cartoon simulation of this AC coupling for a basic approach curve.

Essentially this means the electrical signal from the MCT does not carry the direct distance dependence of the near-field signal anymore since the DC offset present when the tip is closer to the sample is subtracted off. However, the rapidly oscillating parts of the signal still carry direct information about the near-field signal distance dependence. One rising portion of the sinusoidal signal carries direct information about the shape of the near-field distance dependence since it is not skewed –only the DC offset is removed.

The effect of this DC offset only becomes significant when signals are compared for oscillations about points z_0 and z_1 with $\Delta z = z_1 - z_0$ large compared to the distance dependence of the near-field scattering

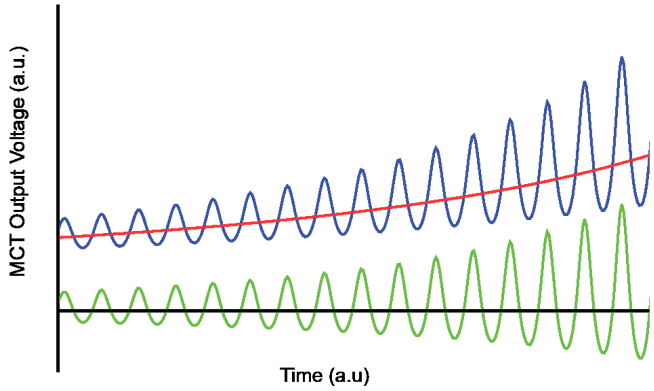


Figure 4.5: This is a cartoon plot of the scattered near-field signal output from the MCT as a function of time as the tip is brought towards the sample at a constant rate. The red curve depicts the signal if the tip is not oscillating. The blue curve represents the signal detected by the MCT as a function of time while the tip is oscillating. The AC filter in the MCT amplifier circuit subtracts off the DC offset (which is approximately equal to the red curve) from the blue curve resulting in the green curve at the output of the MCT.

distance dependence. Thus, if the signal from the MCT and the previously calculated tip-sample separation is correlated with all of the pixels in a given chunk of the data for which Δz between the beginning and end of the chunk is small, this plot will give an accurate reconstruction of that chunk of the raw optical approach curve.

This leads to the idea to locally reconstruct multiple sections of the approach curve and to stitch these chunks together to reconstruct the entire approach curve. From the previous discussion it is clear that each chunk will be centered about the zero line since its DC offset has been subtracted off. To stitch the chunks together, it is necessary to apply some DC offset to each chunk so that they all form one continuous curve. I wrote code which averages various portions of adjacent chunks and aligns them by adding the appropriate DC offsets. Fig. 4.6 shows data before and after this offsetting process.

There are two weaknesses of this measurement which are worth pointing out for future discussion. First, this offset process is somewhat arbitrary. My algorithm offsets the different chunks until they match up so it is no surprise that final result is a continuous curve. However, the smoothness of the curve as well as the decaying behavior indicate that I have, in fact, recreated the actual raw optical approach curve. The second weakness is more fundamental to the real-time acquisition scheme. Since I am looking at the unprocessed

raw signal from the MCT, I am looking at all of the signal falling onto the MCT including the undesirable background signal. The AC coupled MCT eliminates all long range DC components of the background, but many components still persist. Thus my approach curves necessarily contain a large amount of background signal.

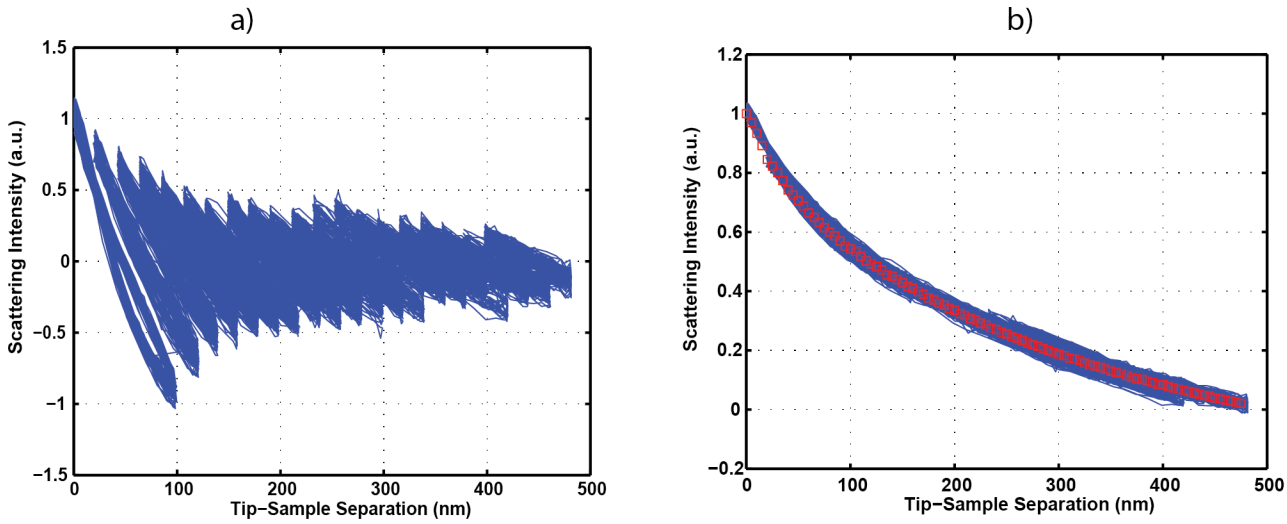


Figure 4.6: a) Detected intensity versus tip-sample separation distance. Each chunk of data is centered about zero because of the DC filtering performed by the MCT amplifier. b) Reconstructed approach curve produced by applying a DC offset to each chunk so that they match up. Red squares represent the average value for a given tip-sample separation.

4.5.2 Results

Fig. 4.7 shows a comparison of the reconstructed raw optical approach curve to harmonic demodulation approach curves. Notice the raw approach initially falls off more shallowly than even the 1st harmonic demodulation but then falls of more steeply. This shows that, as expected, the functional form of the raw approach curve is different from that of the harmonic demodulation. This highlights the point that the harmonic demodulated approach curves represent something different from the distance dependence of the near-field signal and the full analysis process must be considered when interpreting data.

The main motivation for the raw approach curve reconstruction was to compare it to theoretical models for the tip-sample interaction to determine the effectiveness of the model. To this end I fitted the simple dipole model to the reconstructed raw approach curve, however, I found that it was necessary to

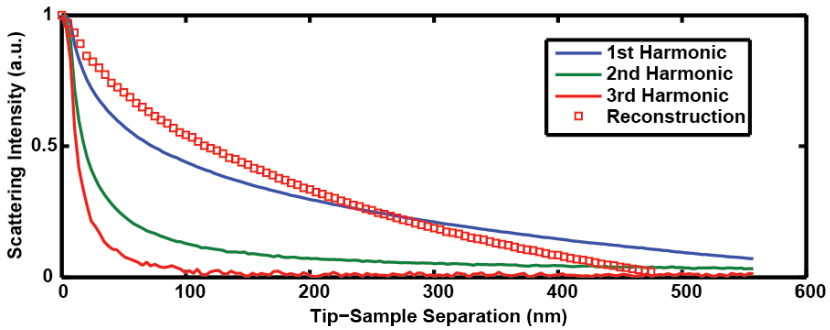


Figure 4.7: The reconstructed raw optical approach curve is compared to simulated harmonic approach curves. Notice the very different functional forms.

include a decaying exponential term so as to model the far-field background present in my signal. I used Mathematica to perform a non-linear model fit. The functional form I attempted to fit is given by equation 2.4. The tip radius was fixed at 100 nm and I used the bulk dielectric functions for platinum and gold for the tip and sample respectively. The free parameters in the fit are characteristic decay length for the exponential, the relative magnitude of the two terms, and a DC offset term.

Fig. 4.8 shows the results of the fitting process. The red curve represents the total fit equation and fits the data extremely well. The green and orange curves represent the tip-sample interaction and far-field background terms. The sum of these two curves plus a DC offset term results in the red curve. For ease of reading, the tip-sample interaction curve has been made to agree with the model when the tip is very close and the far field background has been made to agree when the tip is far away. This plot reveals the relative impact of the near-field interaction and the far-field background on the shape of the total fit curve. At large distances it is clear that the near-field interaction has no effect, but at distances closer than the tip radius of 100 nm, the near-field interaction has a larger relative impact on the shape of the total curve.

This approach curve reconstruction is exciting because it is the first direct reconstruction of the distance dependence of the scattering intensity for an elastic s-SNOM measurement. However, the result thus far is not useful since it is dominated so heavily by the far-field background. Further analysis could be performed on these fits to determine more quantitatively the relative contribution of the near-field signal and the far-field background signal. Because of this large far-field background, it would be difficult at this point to use this reconstruction method to compare the simple-dipole and extended-dipole models since the

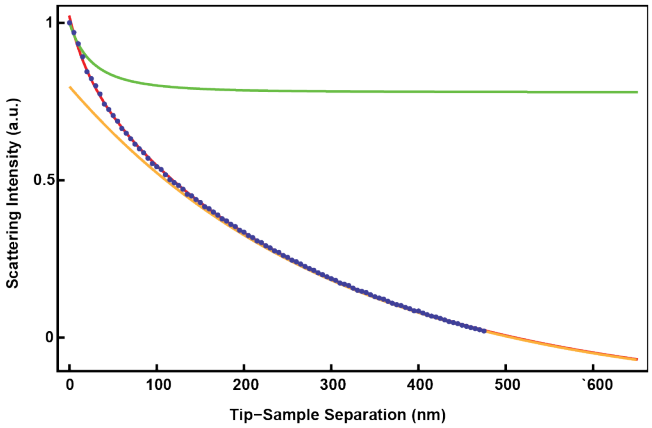


Figure 4.8: Fitting of the raw optical approach curve (blue) by the sum of the simple dipole model (green) and a decaying exponential (orange). Notice the strong agreement between the raw optical approach curve and the total fit function (red). Also notice the relative magnitudes of the near-field interaction and far-field background.

signal is so dominated by the far field background.

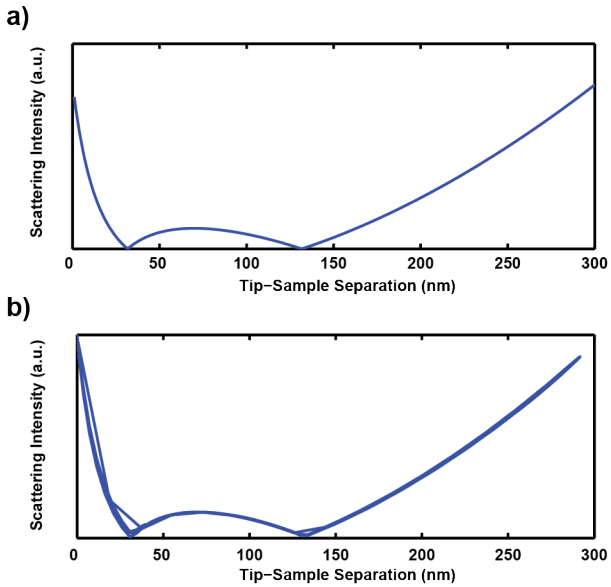


Figure 4.9: Comparison of a) hypothetical scattering distance dependence with b) reconstructed distance dependence using the approach curve reconstruction algorithm. There is good qualitative agreement between the two plots which indicates the algorithm was successful.

To check that the reconstruction process faithfully reconstructs the underlying raw approach curve I tested the algorithm by using a more complex hypothetical scattering intensity distance dependence function. In particular I chose a function which is non-monotonic and which contains discontinuities. My hypothetical

distance dependence of the scattering was $I(z) = \left| \frac{1}{(.01z)^2} + (.01z)^2 - 3 \right|$ shown in Fig. 4.9 a). I simulated the scattered signal as a function of time during an approach curve and applied a high pass AC filter to this signal just like the MCT does to the experimental optical signal. I then used the approach curve reconstruction algorithm on this hypothetical real-time optical data to reconstruct the curve seen in Fig. 4.9 b). This shows that there are not significant artifacts introduced by the reconstruction process.

Chapter 5

Conclusion and Outlook

In this thesis I have exhibited the design and implementation of a new s-SNOM detection scheme based upon real-time acquisition of s-SNOM optical signal and the associated AFM probe positioning signals. I described three techniques to analyze this data, each of which provides insight into s-SNOM itself as a microscopy technique.

I investigated the feasibility of gated-detection s-SNOM. The main result was that simple top and bottom gated illuminated s-SNOM provides the exact same information as would be provided by traditional lock-in detection of the first harmonic. I was able to show that decreasing the tapping amplitude for gated detection provides higher spatial resolution. This is problematic because the first harmonic contains high amounts of background artifact signals. What must be investigated now is if gated detection will still agree strongly with the 1st harmonic lock-in signal for small tapping amplitudes or if the agreement was an effect of the large tapping amplitude.

With this new detection scheme, I was able to reconstruct the distance dependence of the near-field scattering intensity. This reconstruction was made possible by the simultaneous acquisition of both the optical signal and the tip-sample separation. I observed the expected disagreement between the functional forms of the raw approach curve and the harmonic demodulated approach curves.

I also fit the simple-dipole model plus a decaying exponential to the reconstructed approach curve. The fit was very effective and revealed the relative contributions of the near-field interaction and the far-field background to the shape of the raw approach curve. It was clear that the effect of the near-field became much more significant at small tip-sample separations and this is why higher harmonic lock-in

demodulations contain less relative background signal. More analysis could provide quantitative information about the relative contributions of the two signal sources. This technique provides a way to look at the near-field interaction and the background signal separately so it could provide a means to investigate the functional forms of both the tip-sample interaction as well as the far-field background and improve upon models for both.

There are more measurements and simulations required to provide a thorough investigation of this tip-simple interaction. Experimental harmonic approach curves have shown good agreement with theoretical predictions from the simple-dipole and extended-dipole models [38] so it must be investigated if the reconstructed approach curve shows the same agreement.

A self-consistency check would be to take the reconstructed approach curve and, from it, simulate harmonic demodulations. These harmonic demodulations could then be compared to the harmonic demodulations simulated from the raw real-time data. Agreement between these simulations and measurements would imply that there is not some artifact of the reconstruction that is skewing the raw approach curve.

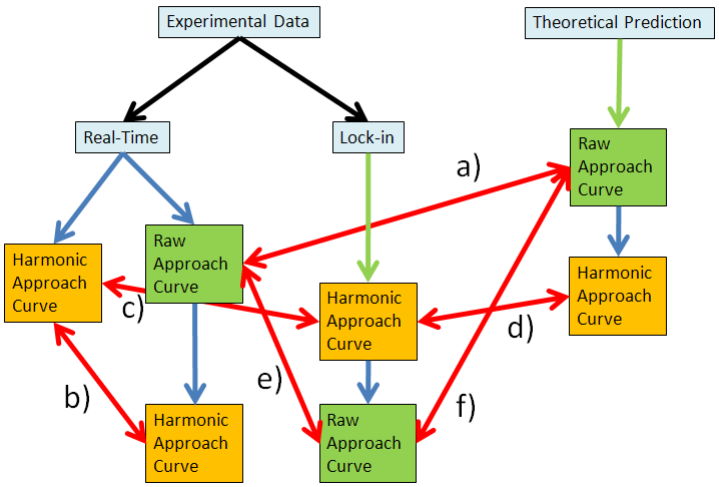
Fig. 5.1 summarizes the measurements and simulations that should be carried out in the future to fully analyze the effectiveness of the new measurement technique. In this network green lines represent immediate results, blue lines represent a simulation step, and red lines represent a comparison of two methods for getting a dataset. Comparison a) represents the work of this thesis. Comparison b) represents the comparison described in the preceding paragraph and would help rule out the worry that the reconstruction is introducing artifacts. Comparison c) could very easily be done on my setup and should be done as a proof of concept –I expect good agreement based on what I've seen in the data. Discrepancies here could reveal important dissimilarities between the two techniques Comparison d) has already seen good agreement in the literature [38][34]. Comparison e) and f) rely on a reconstruction of the raw approach curve from lock-in data. Comparison f) has been made successfully [44] but e) has not been investigated. Comparisons e) and f) show that lock-in data does actually contain all of the real-time information if the lock-in data can be properly 'inverse Fourier transformed.'

This reconstruction technique is very valuable because it could provide a direct means to compare theoretical predictions of the tip-sample interaction to experiment. Further insight into the fundamental

science of this interaction will lead to the improvement of current s-SNOM techniques and also, possibly, the development of new s-SNOM techniques based on new elements of the fundamental understanding. For example, one possibility not predicted by the theoretical models presented here is that the distance dependence of the near-field scattering varies from sample to sample.

From an instrument development standpoint my thesis has been very effective. I have expanded the measurement capabilities of s-SNOM and perhaps more importantly implemented a useful s-SNOM self-diagnostic tool. However, I have also provided a means to study the fundamental science underlying the theory of s-SNOM.

s-SNOM is still a relatively new technique only used by a few research groups. It is important that s-SNOM matures as a measurement technique so that it can allow researchers across many different fields to measure potentially critically important mesoscopic structures in their samples.



Comparison Line	First Dataset	Second Dataset	Note
a)	Recreated raw from real-time	Theoretical raw	Topic of this thesis. Issues explored
b)	Recreated harmonics from real-time	Harmonics from real-time raw reconstruction	Proof that raw approach reconstruction isn't introducing artifacts
c)	Recreated harmonics from real-time	Harmonics from lock-in	Very easy measurement should be done for proof of concept
d)	Harmonics from lock-in	Theoretical harmonics	Explored in earlier literature. We see agreement
e)	Recreated raw from real-time	Recreated raw from lock-in	Proof that it is possible to inverse Fourier transform lock-in data to simulate real time data
f)	Recreated raw from lock-in	Theoretical raw	Same as d), may have same issues as a).

Figure 5.1: Network showing past and future measurements for a complete analysis of the real-time and lock-in data to determine the effectiveness and inter-compatibility of each method. Green lines represent direct results, blue lines represent simulation, and red lines represent necessary comparisons.

Bibliography

[1] J. N. Farahani, D. W. Pohl, H.-J. Eisler, and B. Hecht, Physical Review Letters **95**, 017402 (2005).

[2] a. J. Huber, J. Wittborn, and R. Hillenbrand, Nanotechnology **21**, 235702 (2010).

[3] R. Kasper et al., Small (Weinheim an der Bergstrasse, Germany) **6**, 1379 (2010).

[4] M. B. Raschke et al., Chemphyschem **6**, 2197 (2005).

[5] J. M. Atkin, S. Berweger, A. C. Jones, and M. B. Raschke, Advances in Physics **61**, 745 (2012).

[6] S. a. Centeno, F. Schulte, N. W. Kennedy, and A. G. Schrott, Applied Physics A **105**, 55 (2011).

[7] W. Clauss, D. Bergeron, and a. Johnson, Physical Review B **58**, R4266 (1998).

[8] G. Binning, C. Gerber, and E. Stoll, Surface Science **190**, 1 (1987).

[9] I. Lisiecki, A. Filankembo, and J. Urban, Physical Review B **61**, 4968 (2000).

[10] P. L. Stewart, S. D. Fuller, and R. M. Burnett, THE EMBO Journal **12**, 2589 (1993).

[11] C. Bai, Scanning Tunneling Microscopy and its Applications (Springer Verlag, New York, NY, 2000).

[12] A. Marian, M. C. Stowe, J. R. Lawall, D. Felinto, and J. Ye, Science (New York, N.Y.) **306**, 2063 (2004).

[13] E. Castanié et al., Optics letters **37**, 3006 (2012).

[14] T. Ishikawa et al., Nano letters **12**, 2083 (2012).

[15] G. Saxby, The Science of Imaging (CRC Press, 2002).

[16] R. Wiesendanger, Scanning Probe Microscopy and Spectroscopy: Methods and Applications (Cambridge University Press, 1994).

[17] G. Binning, H. Rohrer, C. Gerber, and E. Weibel, Physical Review Letters **49**, 57 (1982).

[18] G. Binning, C. F. Quate, and C. Gerber, Physical Review Letters **56**, 930 (1986).

[19] K. Kneipp, H. Kneipp, I. Itzkan, R. R. Dasari, and M. S. Feld, Chemical reviews **99**, 2957 (1999).

[20] J. S. Ross et al., Nature communications **4**, 1474 (2013).

[21] S. Berweger et al., Nature nanotechnology **4**, 496 (2009).

[22] G. R. Fowles, Introduction to Modern Optics, Second ed. (Holt, Rinehart and Winston, Inc., New York, NY, 1989).

[23] L. Rayleigh, Philosophical Magazine **8**, 261 (1879).

[24] B. E. A. Saleh and M. Teich, Fundamentals of Photonics (John Wiley & Sons, Inc., New York, NY, 1991).

[25] R. L. Olmon, Optical Vector Near-Field Imaging for the Design of Impedance Matched Optical Antennas and Devices, PhD thesis, University of Washington, 2012.

[26] E. H. Synge, Philosophical Magazine **6**, 356 (1928).

[27] E. A. Ash and G. Nicholls, Nature **237**, 510 (1972).

[28] D. W. Pohl, W. Denk, and M. Lanz, Applied Physics Letters **44**, 70651 (1984).

[29] F. Keilmann and R. Hillenbrand, Philosophical transactions. Series A, Mathematical, physical, and engineering sciences **362**, 787 (2004).

[30] Y. Inouye and S. Kawata, Optics letters **19**, 159 (1994).

[31] N. Behr and M. Raschke, Journal of Physical Chemistry C **112**, 3766 (2008).

[32] P. M. Krenz, R. L. Olmon, B. a. Lail, M. B. Raschke, and G. D. Boreman, Optics express **18**, 21678 (2010).

[33] C. Neacsu, B. van Aken, M. Fiebig, and M. Raschke, Physical Review B **79**, 100107 (2009).

[34] N. Ocelić, Quantitative Near-field, PhD thesis, Technische Universitat Munchen, 2007.

[35] N. Calander and M. Willander, Journal of Applied Physics **92**, 4878 (2002).

[36] A. Cvitkovic, N. Ocelic, and R. Hillenbrand, Optics express **15**, 8550 (2007).

[37] R. Olmon et al., Physical Review Letters **105**, 1 (2010).

[38] M. B. Raschke and C. Lienau, Applied Physics Letters **83**, 5089 (2003).

[39] R. Hillenbrand, B. Knoll, and F. Keilmann, Journal of microscopy **202**, 77 (2001).

[40] T. Ichimura et al., Physical Review Letters **102**, 186101 (2009).

[41] B. D. Mangum, E. Shafran, C. Mu, and J. M. Gerton, Nano letters **9**, 3440 (2009).

[42] N. Ocelic, A. Huber, and R. Hillenbrand, Applied Physics Letters **89**, 101124 (2006).

[43] J. Stadler, C. Stanciu, C. Stupperich, and a. J. Meixner, Optics letters **33**, 681 (2008).

[44] D. Barchiesi and T. Grosges, Optics express **13**, 6519 (2005).